

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ
Федеральное государственное автономное учреждение
высшего профессионального образования
"Казанский (Приволжский) федеральный университет"
Отделение русской и зарубежной филологии



УТВЕРЖДАЮ

Проректор
по образовательной деятельности КФУ
Проф. Таюрский Д.А.

_____ 20__ г.

Программа дисциплины
Компьютерная лингвистика БЗ.ДВ.7

Направление подготовки: 032700.62 - Филология

Профиль подготовки: Прикладная филология: русский язык

Квалификация выпускника: бакалавр

Форма обучения: очное

Язык обучения: русский

Автор(ы):

Каримуллина Г.Н.

Рецензент(ы):

Горобец Е.А.

СОГЛАСОВАНО:

Заведующий(ая) кафедрой: Горобец Е. А.

Протокол заседания кафедры No ____ от "____" _____ 201__ г

Учебно-методическая комиссия Института филологии и межкультурной коммуникации
(отделение русской и зарубежной филологии):

Протокол заседания УМК No ____ от "____" _____ 201__ г

Регистрационный No

Казань
2016

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) доцент, к.н. Каримуллина Г.Н. Кафедра русского языка и прикладной лингвистики отделение русской и зарубежной филологии ,
Guzel.Karimullina@kpfu.ru

1. Цели освоения дисциплины

Дисциплина "Компьютерная лингвистика" посвящена основным понятиям, проблемам, методам и приемам компьютерной лингвистики, изучению основных направлений применения компьютерно-информационных технологий в языкознании.

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

Данная учебная дисциплина включена в раздел " Б3.ДВ.7 Профессиональный" основной образовательной программы 032700.62 Филология и относится к дисциплинам по выбору. Осваивается на 4 курсе, 7 семестр.

Для изучения дисциплины необходимы знания, умения, компетенции, полученные обучающимися в средней общеобразовательной школе, в результате изучения дисциплин математического и естественнонаучного цикла ("Основы лингвографии", "Инновационные технологии в образовании / в науке" и др.) и дисциплин профессионального цикла ("Введение в прикладную филологию").

Изучение данного курса позволяет систематизировать и углубить ранее полученные знания в области прикладной лингвистики.

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

В результате освоения дисциплины формируются следующие компетенции:

Шифр компетенции	Расшифровка приобретаемой компетенции
ОК-7 (общекультурные компетенции)	способностью к самоорганизации и самообразованию (ОК-7);
ОПК-2 (профессиональные компетенции)	способностью демонстрировать знание основных положений и концепций в области общего языкознания, теории и истории основного изучаемого языка (языков), теории коммуникации (ОПК-2);
ОПК-4 (профессиональные компетенции)	владением базовыми навыками сбора и анализа языковых и литературных фактов, филологического анализа и интерпретации текста (ОПК-4);
ОПК-6 (профессиональные компетенции)	способностью решать стандартные задачи профессиональной деятельности на основе информационной и библиографической культуры с применением информационно-коммуникационных технологий и с учетом основных требований информационной безопасности (ОПК-6).
ПК-2 (профессиональные компетенции)	способностью проводить под научным руководством локальные исследования на основе существующих методик в конкретной узкой области филологического знания с формулировкой аргументированных умозаключений и выводов (ПК-2);
ПК-9 (профессиональные компетенции)	владением базовыми навыками доработки и обработки (например, корректура, редактирование, комментирование, реферирование, информационно-словарное описание) различных типов текстов (ПК-9);

В результате освоения дисциплины студент:

1. должен знать:

знать ключевые понятия, связанные с данным направлением прикладной лингвистики;
обладать теоретическими знаниями в области компьютерной лингвистики

2. должен уметь:

ориентироваться в проблемах различных аспектов компьютерной лингвистики

3. должен владеть:

навыками лингвистического поиска в сети Интернет

применять компьютерные технологии на языковом материале

4. Структура и содержание дисциплины/ модуля

Общая трудоемкость дисциплины составляет 3 зачетных(ые) единиц(ы) 108 часа(ов).

Форма промежуточного контроля дисциплины экзамен в 7 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. Компьютерная лингвистика как прикладная лингвистическая дисциплина.	7		2	0	0	
2.	Тема 2. Когнитивный инструментарий компьютерной лингвистики.	7		2	4	0	
3.	Тема 3. Компьютерная лингвография.	7		2	4	0	контрольная работа
4.	Тема 4. Машинный перевод.	7		2	0	0	контрольная работа
5.	Тема 5. Автоматизированные (компьютерные) обучающие системы.	7		0	4	0	

N	Раздел Дисциплины/ Модуля	Семестр	Неделя семестра	Виды и часы аудиторной работы, их трудоемкость (в часах)			Текущие формы контроля
				Лекции	Практические занятия	Лабораторные работы	
6.	Тема 6. Основные понятия корпусной лингвистики	7		2	0	0	
7.	Тема 7. Опыты разработки корпусов текстов.	7		0	8	0	
8.	Тема 8. Корпусы текстов русского языка: история и перспективы развития.	7		0	8	0	письменная работа
9.	Тема 9. Корпусы разговорной речи и аудио-архивы.	7		0	8	0	
	Тема . Итоговая форма контроля	7		0	0	0	экзамен
	Итого			10	36	0	

4.2 Содержание дисциплины

Тема 1. Компьютерная лингвистика как прикладная лингвистическая дисциплина.

лекционное занятие (2 часа(ов)):

Компьютерная лингвистика как прикладная лингвистическая дисциплина. Направления компьютерной лингвистики.

Тема 2. Когнитивный инструментарий компьютерной лингвистики.

лекционное занятие (2 часа(ов)):

Когнитивный инструментарий компьютерной лингвистики. Гипертекстовые технологии представления текста. Понятие гипертекста.

практическое занятие (4 часа(ов)):

Моделирование общения. Компоненты гипертекста. Гипертексты и тексты. Типы гипертекстов.

Тема 3. Компьютерная лингвография.

лекционное занятие (2 часа(ов)):

Аккумулятивный аспект компьютерной лингвистики. Компьютерная лингвография.

практическое занятие (4 часа(ов)):

Компьютерный комплексный языковой справочник и его свойства. Области применения комплексного языкового справочника.

Тема 4. Машинный перевод.

лекционное занятие (2 часа(ов)):

Машинный перевод. История машинного перевода. Основные проблемы машинного перевода. Современные системы машинного перевода. Области использования машинного перевода.

Тема 5. Автоматизированные (компьютерные) обучающие системы.

практическое занятие (4 часа(ов)):

Автоматизированные (компьютерные) обучающие системы. Основные компоненты компьютерных обучающих систем. Функции компьютерных обучающих систем.

Тема 6. Основные понятия корпусной лингвистики

лекционное занятие (2 часа(ов)):

Основные понятия корпусной лингвистики: проблемная область, корпус данных и корпус текстов, единица хранения корпуса данных / текстов. Типология корпусов текстов: исследовательские корпуса, иллюстративные корпуса, статистические и динамические (мониторные) корпуса, синхронические и диахронические корпуса и др. Способы представления и хранения корпуса данных. Требования к корпусу текстов с точки зрения пользователя (репрезентативность, полнота, структурность, экономичность и др.).

Тема 7. Опыты разработки корпусов текстов.

практическое занятие (8 часа(ов)):

Опыты разработки корпусов текстов. Фундаментальные корпуса английского, французского, немецкого, китайского и др. языков.

Тема 8. Корпусы текстов русского языка: история и перспективы развития.

практическое занятие (8 часа(ов)):

Корпусы текстов русского языка: история и перспективы развития. Тюбингенский корпус русского языка. Хельсинкский аннотированный корпус (ХАНКО). Корпус газетных текстов русского языка конца XX века (КГТ). Национальный корпус русского языка (НКРЯ). Структура корпуса. Состояние и перспективы развития НКРЯ.

Тема 9. Корпусы разговорной речи и аудио-архивы.

практическое занятие (8 часа(ов)):

Корпусы разговорной речи и аудио-архивы. Корпусы русской разговорной речи. Электронные библиотеки. Электронные библиотеки и корпуса текстов: сходства и различия. Основные электронные библиотеки русских текстов

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
3.	Тема 3. Компьютерная лингвография.	7		подготовка к контрольной работе	16	контрольная работа
4.	Тема 4. Машинный перевод.	7		подготовка к контрольной работе	16	контрольная работа
8.	Тема 8. Корпусы текстов русского языка: история и перспективы развития.	7		подготовка к письменной работе	12	письменная работа
	Итого				44	

5. Образовательные технологии, включая интерактивные формы обучения

Лекции, практические занятия, самостоятельная работа студентов.

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

Тема 1. Компьютерная лингвистика как прикладная лингвистическая дисциплина.

Тема 2. Когнитивный инструментарий компьютерной лингвистики.

Тема 3. Компьютерная лингвография.

контрольная работа , примерные вопросы:

1. Редактирование словарных статей происходит а. в обычном текстовом файле б. в базе данных в. в специально созданной программе 2. Автоматический словарь ? это а. словарь в специальном машинном формате, предназначенный для использования на ЭВМ пользователем или компьютерной программой обработки текста б. компьютерная форма представления словарной информации и целевой ориентацией в. справочник, содержащий языковые элементы, расположенные в определенном порядке, объясняет значения описываемых единиц, дает различную информацию о них или их перевод на другой язык либо сообщает сведения о предметах, обозначаемых ими 3. Сколько видов упорядочения материалов существует в КЯС? а. 1 б. 2 в. несколько 4. Какая именно база лингвистических данных носит лингвографический характер? а. разработки, содержащие в названии термины словарь или справочник б. представленная в электронном варианте в. любая 5. Что из перечисленного относится к инвентариям? а. буква, слог, словоформа и т.п. б. слово[формо]указатели, индексы и т.п. в. выделение полужирным шрифтом, курсивом, цветом и т.п. 6. Чему КЯС не может противопоставляться? а. некомплексным словарям б. переводным словарям в. неавтоматизированным словарям 7. Какие словари из перечисленных относятся к однопараметровым ЯС? а. словари фразеологизмов б. морфологические в. орфографические 8. Какие из перечисленных ЯС способны выполнять обучающую функцию? а. традиционные б. автоматизированные в. любые

Тема 4. Машинный перевод.

контрольная работа , примерные вопросы:

Машинный перевод 1. Начало работ по машинному переводу относят к а) началу 80-х гг. XX века б) середине 60-х гг. XX века в) 50-м гг. XX века 2. Первая международная конференция по МП прошла в а) Америке б) Австралии в) Германии г) Японии 3. Джорджтаунский эксперимент связан с переводом а) с русского языка на английский б) с русского языка на немецкий в) с английского языка на русский г) с немецкого языка на русский 4. Словарь программы для Джорджтаунского эксперимента включал а) 3500 слов б) 500 слов в) 250 слов 5. Первоначально системы МП должны были затронуть а) художественные тексты б) научные тексты в) публицистические тексты г) официально-деловые тексты д) все выше перечисленные тексты 6. Почему, несмотря на неудачи в области МП, некоторые заказчики продолжали использовать данный вид перевода: а) их мало интересовало качество перевода б) по привычке в) их привлекал объем выдаваемой продукции г) их устраивали небольшие финансовые затраты д) из-за всех выше перечисленных причин 7. Первые системы МП характеризуются стратегией а) ограниченного машинного перевода б) прямого перевода в) простого перевода 8. При стратегии перевода через язык-посредник а) анализ ведется в категориях входного языка, синтез ? в категориях выходного б) анализ ведется в категориях выходного языка, синтез ? в категориях входного в) в каких категориях ведется анализ и синтез не имеет значения

Тема 5. Автоматизированные (компьютерные) обучающие системы.

Тема 6. Основные понятия корпусной лингвистики

Тема 7. Опыты разработки корпусов текстов.

Тема 8. Корпусы текстов русского языка: история и перспективы развития.

письменная работа , примерные вопросы:

Схема описания корпуса 1. Время создания корпуса. 2. Цель корпуса. 3. Создатели корпуса (участники проекта). 4. Объем корпуса. 5. Источники (материалы) корпуса (тексты какого жанра, какого периода и др.). 6. ?Опора? корпуса (исследования, словари, использованные при создании). 7. Предназначение корпуса. 8. Структура корпуса (краткая характеристика разделов, подкорпусов). 9. Тип корпуса (подкорпусов) (см. классификацию). 10. Типы представленных в корпусе разметок, их характеристика. 11. Организация поиска в корпусе. 12. Что создано на основе материалов данного корпуса (справочники, грамматики и т.п.). 13. Преимущества корпуса. 14. Недостатки корпуса. Корпусы для анализа ? Национальный корпус русского языка ? Машинный фонд русского языка ? Электронная библиотека русских народных говоров (<http://dialekt.rx5.ru/>)

Тема 9. Корпусы разговорной речи и аудио-архивы.

Тема . Итоговая форма контроля

Примерные вопросы к экзамену:

Вопросы к экзамену.

1. Компьютерная лингвистика как прикладная лингвистическая дисциплина. Основные направления компьютерной лингвистики.
2. Метаязык компьютерной лингвистики (фрейм, сценарий, план, модель мира).
3. Моделирование общения как направление компьютерной лингвистики.
4. Моделирование сюжета как направление компьютерной лингвистики.
5. Машинный перевод. Основные этапы развития машинного перевода.
6. Причины невозможности использования искусственного интеллекта для перевода любых текстов (художественных, например).
7. Классификация систем машинного перевода.
8. Области использования машинного перевода.
9. В чем преимущество компьютерных словарей?
10. Сферы использования компьютерных справочников.
11. Обработка естественного языка как направление компьютерной лингвистики.
12. Жесткая структура диалога, мягкая структура диалога.
13. Архитектура систем обработки естественного языка.
14. Компьютерная лингвография: предмет и задачи.
15. Современные системы машинного перевода.
16. Звуковые базы данных и аудио-архивы.
17. Звуковые базы данных русского языка.
18. Подкорпуса Национального корпуса русского языка
19. "Машинный фонд русского языка".
20. Национальный корпус русского языка: особенности и перспективы развития.
21. Основные свойства корпуса текста.
22. Корпусы текстов английского языка.
23. Корпусы текстов немецкого языка
24. Типология корпусов текстов.
25. Разметка. Типы разметок.

7.1. Основная литература:

Информатика, автоматизированные информационные технологии и системы: Учебник / В.А. Гвоздева. - М.: ИД ФОРУМ: ИНФРА-М, 2011. - 544 с.

[//http://www.znaniyum.com/bookread.php?book=207105](http://www.znaniyum.com/bookread.php?book=207105)Бовтенко М.А. Компьютерная лингводидактика: учеб.пособие / М.А.Бовтенко. - М.: Флинта: Наука, 2005. - 215 с. (3 экз.)

Захарова И.Г. Информационные технологии в образовании: учебное пособие для студентов / И.Г. Захарова. - 4-е изд., стер. - М.: Академия, 2008. - 187 с. (5 экз.)

7.2. Дополнительная литература:

Боженюк А.В. Интеллектуальные интернет-технологии: учебник / А.В. Боженюк, Э.М. Котов, А.А. Целых. - Ростов-на-Дону: Феникс, 2009. - 381 с. (20 экз.)

Зубов А.В. Информационные технологии в лингвистике: учеб.пособие / А.В.Зубов, И.И.Зубова. - М.: Academia, 2004. - 205 с. (31 экз.)

Информационные технологии в профессиональной деятельности: Учебник / В.Н. Гришин, Е.Е. Панфилова. - М.: ИД ФОРУМ: НИЦ ИНФРА-М, 2013. - 416 с. //

<http://znaniyum.com/bookread.php?book=3989121>.

Информационные технологии в профессиональной деятельности: Учебное пособие / Е.Л. Федотова. - М.: ИД ФОРУМ: НИЦ Инфра-М, 2012. - 368 с. // <http://znanium.com/bookread.php?book=322029>

7.3. Интернет-ресурсы:

Русский филологический портал - www.gramota.ru

Сайт лаборатории компьютерной лингвистики - <http://proling.iitp.ru/>

Сайт Машинного фонда русского языка - <http://cfri.ru/>

Сайт Национального корпуса русского языка - www.ruscorpora.ru

Сайт по компьютерной лингвистике - <http://vp-zakharov.narod.ru>

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.

Учебно-методическая литература для данной дисциплины имеется в наличии в электронно-библиотечной системе "БиблиоРоссика", доступ к которой предоставлен студентам. В ЭБС "БиблиоРоссика" представлены коллекции актуальной научной и учебной литературы по гуманитарным наукам, включающие в себя публикации ведущих российских издательств гуманитарной литературы, издания на английском языке ведущих американских и европейских издательств, а также редкие и малотиражные издания российских региональных вузов. ЭБС "БиблиоРоссика" обеспечивает широкий законный доступ к необходимым для образовательного процесса изданиям с использованием инновационных технологий и соответствует всем требованиям федеральных государственных образовательных стандартов высшего профессионального образования (ФГОС ВПО) нового поколения.

Учебно-методическая литература для данной дисциплины имеется в наличии в электронно-библиотечной системе "ZNANIUM.COM", доступ к которой предоставлен студентам. 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, УМК, 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 профессионального образования (ФГОС ВПО) нового поколения.

Компьютерный класс с доступом к сети Интернет, проектор, экран, учебная доска.

Программа составлена в соответствии с требованиями ФГОС ВПО и учебным планом по направлению 032700.62 "Филология" и профилю подготовки Прикладная филология: русский язык .

Автор(ы):

Каримуллина Г.Н. _____

"__" _____ 201__ г.

Рецензент(ы):

Горобец Е.А. _____

"__" _____ 201__ г.