ТИПОВАЯ ПРОГРАММА
по дисциплине «РЕЧЕВАЯ ИНФОРМАТИКА»

для направления подготовки 531200 «Компьютерная лингвистика»
Академическая степень: бакалавр

Составитель: **Кожалиева Д. Ж.** - преп. кафедры Информационных технологий факультета Лингвистики

СТРУКТУРА ПРОГРАММЫ

I. Организационно-методический раздел
- I. 1. Цели и задачи дисциплины
- I. 2. Место дисциплины в структуре ООП высшего профессионального образования
- I. 3. Требования к результатом освоения содержания дисциплины
- I. 4. Объем дисциплины и виды учебной работы
- I. 5. Формы контроля
- I. 6. Технологическая карта дисциплины

II. Содержание программы учебной дисциплины
- II. 1 Содержание разделов дисциплины
- II. 2 Вопросы и задания для самостоятельной работы студентов

III. Оценочные средства по итогам освоения дисциплины
- III. 1 Примерные темы проектов
- III. 2 Примерные вопросы к экзамену

IV. Образовательные технологии
- IV. 1 Учебно-методические: основная и дополнительная литература
- IV. 2 Информационное и материально-техническое обеспечение дисциплины: база-данных, информационно-справочные и поисковые системы
- IV. 3 Материально-техническое обеспечение дисциплины

I. ОРГАНИЗАЦИОННО-МЕТОДИЧЕСКИЙ РАЗДЕЛ

I.1. Целью дисциплины «Речевая информатика» является освоение студентами фундаментальных знаний в области обработки и анализа речевой информации, изучение основных проблем компьютерной обработки речи современных подходов к их решению.

Задачи курса:

- формирование базовых знаний в области компьютерной обработки речевой и аудио информации как дисциплины, обеспечивающей технологические основы современных инновационных сфер деятельности;
- обучение студентов принципам решения задач обработки (цифровой обработки, анализа, распознавания) естественной речи на основе методов машинного обучения и распознавания образов;
- подготовка студентов к выполнению самостоятельных исследований в области речевой технологии.

Пререквизиты – это дисциплины, содержащие знания, умения и навыки, необходимые для освоения изучаемой дисциплины, соответственно до изучения данной дисциплины определяют ее преемственность. Пререквизитами данного курса являются знания студентов по информатике.

Постреквизиты– это дисциплины, для изучения которых требуются знания, умения и навыки, приобретаемые по завершении 1-го курса, т. е. это обязательные для освоения дисциплины, соответственно после изучения данного предмета. Постреквизитами данного курса являются практическое применение сетевых и информационных технологий, а также теоретические дисциплины согласно учебному плану.

I.2. Место дисциплины в структуре ООП бакалавриата

Дисциплина «Речевая информатика» относится к дисциплинам базовой части профессионального цикла учебного плана подготовки студентов по направлению 531200 “Компьютерная лингвистика”.

Для успешного изучения курса студенту необходимо знать общесистемное программное и техническое обеспечении автоматизированных систем, а также уметь работать с персональной ЭВМ.

Курс содержит теоретическую и практическую части

Лекционный курс должен дать базовые знания о компьютерных технологиях, общие понятия информационных технологий, информационных систем и баз данных, дать характеристику интеллектуальных систем, понятия о лингвистических сетевых технологиях и системах.

Курс практических занятий спланирован следующим образом: введение новых понятий и методов работы с ними (для этого можно применить теоретический лекционный курс) выполнение заданий для получения основных навыков работы (лабораторные работы и упражнения) выполнение заданий самостоятельной работы.

1.3. Требования к результатам освоения содержания дисциплины

Компетенции бакалавра, формируемые в результате освоения дисциплины:

-владение культурой мышления, способность к обобщению, анализу, восприятия информации, постановке цели и выбору путей её достижения, умение логически верш аргументировано и ясно строить устную речь;

-владение широкой общей подготовкой (базовыми знаниями) для решения практических задач в области информационных систем и технологий;

-способность самостоятельно работать на компьютере (профессиональные навыки);

-владение навыками использования программных средств и работы в компьютерных сетях, использования ресурсов Интернет; владение основными методами, способами и средствами получения, хранения, переработки информации;

-готовность работать с информацией из различных источников;

-умение критически оценивать свои достоинства и недостатки, наметить пути и выбрать средства развития достоинств и устранения недостатков;

-способность создавать системы автоматической рубрикации текстов, а также информационно-поисковые системы, основанные на разных моделях поиска;

-способностью разрабатывать и внедрять в практику компьютерные системы обучения;

-способность осуществлять организацию рабочих мест, их техническое оснащение, размещение компьютерного оборудования;

В результате освоения компетенций студент должен:

Знать:

- знать системы разработки программ с использованием компьютерной технологии;

- знать методы и уровни представления данных, способы обработки и хранения данных;

Уметь:

- уметь ориентироваться в современных логических языках и возможностях ;

- уметь обосновывать выбор языка для решения конкретных задач;

Владеть:

- информацией о современных автоматизированных системах, используемых для распознания речи;

I.4. Объем дисциплины и виды учебной работы для 2-курса

Виды учебной работы	**Всего**		**Семестры**	
	Часов	**кредиты**	**3**	**4**
Аудиторные занятия (всего)	76	5	36	40
В том числе				
Лекционные занятия	30		16	14
Практические занятия	**46**		**22**	**23**
Самостоятельная работа (всего)	**76**		**36**	**40**
Вид промежуточной аттестации				**Экзамен**
Общая трудоемкость часов	**152**			

I. 5. Формы контроля: экзамен

Текущий контроль осуществляется лектором и преподавателем, ведущим практические занятия.

Контроль за освоением дисциплины осуществляется в каждом дисциплинарном разделе отдельно.

Рубежный контроль (модуль): устный опрос по отдельным разделам дисциплины, выполнение практических работ.

Рейтинговая оценка знаний обучающихся

Формой промежуточной аттестации по дисциплине «Речевая информатика» является контрольная работа или зачет. При этом система текущего и рубежного контроля строится по балльно-рейтинговой модели, а в результате итогового контроля (экзамен) студенты получают в качестве накопительной рейтинговой оценки (не менее 60 баллов.)

Экзамены по дисциплине «Речевая информатика» на 2 курсе проводятся в соответствии с учебными планами.

В технологической карте выделены все формы текущего и рубежного контроля, каждая из которых оценивается определенным количеством баллов.

Текущий контроль осуществляется:

а) путем выполнения студентами вопросов для самопроверки

б) путем выслушивания ответов на заданные вопросы, беседы (преподаватель-студент и студент-студент), интерактивная работа с помо-

щью упражнений, грамматических заданий, экспертная оценка части практических занятий, представленная студентами.

Рубежный контроль осуществляется в виде написания самостоятельных и контрольных работ, как по прохождении разделов курса, так и по прохождении тем внутри каждого модуля.

Технологическая карта включает базовую и дополнительную части. В базовой части приводиться расчет рейтинговой оценки

Тех учебных процедур, которые выполняются студентами в течение семестра. Для итогового экзамена студентам достаточно накопить 60 баллов.

Дополнительная часть технологической карты включает компенсирующие учебные задания. Студенты выполняют их в тех случаях, если в рамках базовой части они не сумели набрать количество баллов, необходимое для получения экзамена. Одним из таких заданий является написания сбора научной информации по теме, предложенной преподавателем (в рамках проблематики курса); написание творческой работы по заданной теме, сбор лексикографического материала (создание глоссария, дефиниции терминов).

Выбор дополнительных учебных заданий из предложенного перечня студенты выполняют самостоятельно. Срок их выполнения определяется последней учебной неделей семестра.

Контроль знаний усвоения: Успешное формирование знаний и умений зависит от систематического контроля успеваемости студентов.

Повседневный контроль: проверка упражнений, знание выражений, грамматики и лексики.

Промежуточный контроль: проверка заданий, выполненных самостоятельно в аудитории или дом, устный и письменный контроль лексики и выражений.

Итоговый контроль: модули, зачеты и экзамены.

I.6. Технологическая карта дисциплины

Наименование дисциплины	Уровень/ступень образования (бакалавриат/ магистратура)	Количество зачетных единиц/кредитов	Форма отчетности	Курс, семестр
Речевая информатика	бакалавриат	5	Экзамен	2 курс 3-4семестр

БАЗОВЫЙ МОДУЛЬ (проверка знаний и умений по дисциплине)				
Тема или задание текущей аттестационной работы	**Виды текущей аттестации**	**Аудиторная или внеаудиторная**	**Минимальное количество баллов**	**Максимальное количество баллов**
Выполнение заданий в течение учебного процесса	**Своевременное выполнение домашних и проектных работ**	**Внеаудиторная или аудиторная**	**30**	**40**
Выполнение заданий самостоятельную работу(Написание реферата, проектных работ по задан. теме)	**Выполнение определенного задание по СРС**	**внеаудиторная**	**5**	**15**
Знание базовых понятий и примеров по основным темам и умение решить практических задач	**Контрольная работа, тестирование**	**Аудиторная и внеаудиторная**	**25**	**30**
Итого:			**60**	**85**

ДОПОЛНИТЕЛЬНЫЙ МОДУЛЬ				
Тема или задание текущей аттестационной работы	**Виды текущей аттестации**	**Аудиторная и внеаудиторная**	**Минимальное количество баллов**	**Максимальное количество баллов**
Проект, презентация	Устный ответ	Аудиторная	4	10
Посещение занятий	Присутствие	Аудиторная	1	5
Итого:			5	15
Итого всего:			65	100

Все оценки складываются из результатов накопленных баллов по каждому модулю и для финального контроля выводится средний балл для окончательной оценки за курс:

Шкала перевода баллов в оценки следующее:

1 модуль -70 баллов	СРС-15 баллов
0-51 «неуд»	0-3 «неуд»
52-62 «удовл»	4-8 «удовл»
63-74 «хор»	9-11 «хор»
75-85 «отл»	12-15 «отл»

Проверка знаний: 3 контрольные работы (или тестирование) по основным темам и 2 проектные работы.

Дополнительные требования для студентов, отсутствующих на занятиях по уважительной причине: устное сквозное собеседование с преподавателем по проблеме пропущенных практических занятий, обязательное выполнение аудиторных проверочных и письменных работ.

Итоговая форма контроля – экзамен.

II. СОДЕРЖАНИЕ ПРОГРАММЫ УЧЕБНОГО КУРСА

II.1. Содержание разделов дисциплины

Предмет речевая информатика. Основные задачи речевой технологии.

Действия с информацией. Сбор информации. Представление информации. Кодирование информации. Декодирование информации. Хранение информации. Обработка информации.

Объект и его характеристика. Объект. Имя объекта. Свойства объекта. Общие и отличительные свойства. Существенные свойства и принятие решения. Элементный состав объекта. Действия объекта. Отношения между объектами.

Информационный объект и компьютер. Информационный объект и смысл. Документ как информационный объект. Электронный документ и файл. Текст и текстовой редактор. Изображение и графический редактор. Схема и карта. Число и электронные таблицы.

Автоматическая обработка звучащей речи.

Ментальное и физическое представление речевого сигнала: основные характеристики и принципиальные различия. Характер связи между ментальным и физическим представлениями.

Основные методы автоматического распознавания речи.

Динамическое программирование, скрытые марковские цепи. Подробная характеристика одного из этих методов.

Основные методы автоматического синтеза речи.

Формантный синтез, КЛП-синтез, компилятивный синтез и их современные модификации.

Лингвистический подход к автоматическому распознаванию речи. Метод чтения динамических спектрограмм как основа построения базы знаний (экспертная система) для лингвистического декодирования речевых сигналов.

Лингвистическое обеспечение для автоматического синтеза речи. Алгоритмы транскрибирования и кодировки текстов и аллофонная база данных.

Примерный план практических занятий

1. Кодирование информации. Декодирование информации. Хранение информации. Обработка информации.
2. Информационный объект и компьютер. Информационный объект и смысл.
3. Электронный документ и файл.
4. Текст и текстовой редактор
5. Редактирования и корректирования текстов
6. Изображение и графический редактор. Схема и карта.
7. Число и электронные таблицы.
8. Звуковые редакторы
9. Использование анимации
10. Синтезаторы речи.
11. Автоматизированный перевод

II.2. Вопросы и задания для самостоятельной работы студентов

1. Речевая информатика
2. Что такое информация
3. Объект и его характеристика.
4. Свойства объекта
5. Текст и текстовой редактор
6. Системы автоматизированной обработка текста.
7. Синтез речи
8. Лингвистическоеобеспечениеавтоматическогосинтезаречи
9. Основы речевой коммуникации
10. Какова архитектура одной из систем автоматизированной обработки текста?
11. Какие формальные языки представления текста вы знаете?
12. Каким образом реализуется автоматизация синтеза текста?

13. Каким образом реализуется автоматизация обработки текста в текстовом процессоре?
14. Какие есть способы и инструменты автоматической обработки текста в текстовом процессоре?
15. Что такое тезаурус
16. Словари в текстовом процессоре?
17. Основные методы автоматического распознавания речи.

III. ОЦЕНОЧНЫЕ СРЕДСТВА ПО ИТОГАМ ОСВОЕНИЯ ДИСЦИПЛИНЫ

III.1. Примерные вопросы к экзамену

1. Предмет речевая информатика.
2. Основные задачи речевой технологии.
3. Сбор информации. Представление информации.
4. Кодирование информации. Декодирование информации.
5. Хранение информации. Обработка информации.
6. Объект и его характеристика.
7. Объект. Имя объекта.
8. Свойства объекта. Общие и отличительные свойства. Существенные свойства и принятие решения.
9. Информационный объект и компьютер. Информационный объект и смысл. Документ как информационный объект. Электронный документ и файл.
10. Текст и текстовой редактор. Изображение и графический редактор. Схема и карта. Число и электронные таблицы.
11. Ментальное и физическое представление речевого сигнала: основные характеристики и принципиальные различия.
12. Динамическое программирование, скрытые марковские цепи.
13. Основные методы автоматического синтеза речи.
14. Формантный синтез, КЛП-синтез, компилятивный синтез и их современные модификации.
15. Лингвистический подход к автоматическому распознаванию речи. Метод чтения динамических спектрограмм как основа построения базы знаний (экспертная система) для лингвистического декодирования речевых сигналов.
16. Лингвистическое обеспечение для автоматического синтеза речи.
17. Алгоритмы транскрибирования и кодировки текстов и аллофонная база данных.

IV. ОБРАЗОВАТЕЛЬНЫЕ ТЕХНОЛОГИИ

IV.1. Учебно-методические: основная и дополнительная литература

Основная литература:

1. Ронжин А. Л., Карпов А. А., Ли И. В. Речевой и многомодальный интерфейсы, Наука, 2006 г.
2. Дж. Кейтер. Компьютеры – синтезаторы речи. М., Мир, 1985.
3. *Н. К. Обжелян, В. Н. Трунин-Донской.* Речевое общение в системах "человек-ЭВМ". Кишинев, "Штиинца", 1985.
4. Потапова Р. К. Речевое управление роботом: лингвистика и современные автоматизированные системы / Р. К. Потапова. - М.: Ком Книга, 2005. - 328 с.
5. Винцюк Т. К. Анализ, распознавание и смысловая интерпретация речевых сигналов, Наукова думка, 1987
6. Рабинер Л, Шафер Р. Цифровая обработка речевых сигналов, М., Радио и связь, 1981
7. Маркел Дж., Грей. Линейное предсказание речи, М., Радио и связь, 1980
8. Методы автоматического распознавания речи под ред. У. Ли, М. Мир 1983 г.
9. Corpus-Based Methods in Language and speech processing, Kluwer Academic Publishers, 1997[библиотекаВЦРАН]
10. X. Huang, AceroA. Spoken Language Processing: a Guide to Theory, Algorithm and System Development, 2001, Prentice Hall [доступна в http: //www. ccas. ru/sites/speech]
11. Rabiner L., Juang B. -H. Fundamentals of Speech Recognition. Prentice Hall, 1993 Prentice Hall PTR[доступнавhttp: //www. ccas. ru/sites/speech]

IV.2. Информационное и материально-техническое обеспечение дисциплины.

1. Proceedings of IEEE Conference on Acoustic, Speech and Signal Processing (1980-2012)
2. Proceedings of InterSpeechConference(1995-2012 гг.)
3. Proceedings of SPECOM (Speech and ComputerInt. Conference)(1998-2012 гг.)
4. Speech Communication, Elsevier
5. Computer Speech and Language, Academic Press, Elsevier

6. IEEE Transactions on Audio, Speech and Language Processing (IEEE Signal Processing Society)
7. Речевые технологии, м., Нар. образование (2008-2012)
8. htk. eng. cam. ac. uk/- HTKToolkit- сайт Кембриджский университета http: //speech. cs. cmu. edu/ сайт университета Карнеги-Меллона, США

IV.3. Материально-техническое обеспечение дисциплины

Необходимо наличие компьютерного класса со следующим программным обеспечением:

Необходимое оборудование для лекций и практических занятий: Компьютер и мультимедийное оборудование (проектор, звуковая система)

Операционные системы: WindowsXP или Windows 7;

Прикладные программные средства:
Microsoft Office, Power Point, MSWindows,
браузер InternetExplorer 6. 0 или выше;
браузер MozillaFireFox 5. 0 или выше.

Обеспечение самостоятельной работы - базы данных по журналам ComputationalLinguistics, ACL Journal