ТИПОВАЯ ПРОГРАММА
По дисциплине «АВТОМАТИЗИРОВАННАЯ ОБРАБОТКА ЕСТЕСТВЕННОГО ЯЗЫКА»

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

Составитель: **Мадраимова Ж. И.** – и. о. доц. кафедры Информационных технологий факультета Лингвистики

СТРУКТУРА ПРОГРАММЫ

1. Организационно-методический раздел
1.1 Цели и задачи дисциплины
1.2 Место дисциплины в структуре ООП высшего профессионального образования
1.3 Требования к результатом освоения содержания дисциплины
1.4 Объем дисциплины и виды учебной работы
1.5 Формы контроля
1.6 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. Цель дисциплины: обусловлена необходимостью решения задачи разработки алгоритмов для проведения достоверного автоматического анализа постоянно увеличивающегося объема информации, порождаемого в рамках информационно-коммуникационных сред

Задача дисциплины:

- проведение анализа существующих систем автоматической обработки текстовой информации,
- разработка поэтапного алгоритма автоматического анализа естественно-языковых текстов,
- создание модели извлечения смысла посредством автоматического анализа и синтеза информационного корпуса.

Пререквизиты: знать фундаментальные основы информатики, базовые концепции технологий программирования и теоретические основы современных информационных систем.

Постреквизиты: общие принципы построения систем автоматической обработки текста (письменного и устного), этапы и уровни автоматической обработки текста.

I.2. Место дисциплины в структуре ООП

Данная дисциплина относиться к циклу профессиональных дисциплин учебного плана по направлению «Компьютерная лингвистика»

Курс «Автоматическая обработка естественного языка» опирается на знания, умения и навыки, сформированные при изучении предмета «Информационные технологии в лингвистике», «Компьютерная лингвистика» в общеобразовательных учебных заведениях.

I.3. Требования к результатам освоения содержания дисциплины

Данный курс является специализированным и дает студентам возможность расширить знания в индустрии информационно-поисковых систем, развитие которых стимулированно расширением сферы Интернета, сосуществует с фактором относительно слабой развитости автоматизированных средств анализа естественно языковой информации, что связано с недостаточной степенью разработки соответствующего лингвистического обеспечения.

В результате освоения дисциплины студент должен овладеть следующими компетенциями:

1) общенаучными (ОК):

- понимание социальной значимости своей будущей профессии, обладание высоко мотивацией к выполнению профессиональной дея-

тельности (ОК- 3);

- владение широкой общей подготовкой (базовыми знаниями) для решения практических задач в области информационных систем и технологий (ОК-6);

2) инструментальными (ИК):

- способность самостоятельно работать на компьютере (профессиональные навыки) (ИК-1);
- владение навыками использования программных средств и работы в компьютерных сетях, использования ресурсов Интернет; владение основными методами, способами и средствами получения, хранения, переработки информации;
- способность использовать современную вычислительную технику и специализированное программное обеспечение в научно-исследовательской работе (ИК-6);

3) социально-личностными и общекультурными (СЛК);

- способность проектировать и осуществлять индивидуально-личностные концепции профессиональной деятельности (СЛК-6);
- готов к самооценке, ценностному социокультурному самоопределению и саморазвитию (СЛК-8);'
- готов к позитивному, доброжелательному стилю общения (СЛК-9);

4) профессиональными (ПК);

- способность разрабатывать алгоритмы частичного синтаксического анализа текстов, написанных на естественном языке (ПК-1);
- способность использовать морфосинтаксические свойства предложений и применять современные технологии морфосинтаксического анализа для решения задач обработки текстов (ПК-2);
- Производственно-технологическая деятельность:
- способность создавать системы автоматической рубрикации текстов, а также информационно-поисковые системы, основанные на разных моделях поиска (ПК-8);
- способность разрабатывать программное обеспечение для автоматизации формирования и пополнения лингвистических ресурсов (ПК-9);
- способность применять современные методы автоматической обработки текстов, статистики, машинного обучения для извлечения знаний из текстов (ПК-10);

Производственно-практическая деятельность:

- способностью разрабатывать и тестировать лингвистические процессоры (ПК-
- готовностью к созданию лингвистических компонентов электрон-

ных языковых ресурсов (корпусов текстов, словарей, фонетических, лексических, терминологических баз данных) (ПК-21);

- способностью разрабатывать лингвистические компоненты систем автоматической обработки естественного языка (синтеза и распознавания устной речи, генерации текста, контент-анализа, автоматического перевода, автоматического реферирования и аннотирования), а также интеллектуальных систем (вопросно- ответных, экспертных) (ПК-22);
- способность разрабатывать лингвистические компоненты систем автоматической обработки естественного языка (синтеза и распознавания устной речи, генерации текста, контент-анализа, автоматического перевода, автоматического реферирования и аннотирования), а также интеллектуальных систем (вопросно- ответных, экспертных) (ПК-28);

В результате освоения дисциплины «Автоматическая обработка естественного языка»студент должен:

знать:

- методы и решения в системах организации знаний; эмпирические законы описания естественных языков; логико-статистические методы анализа языкового описания предметных областей; технологию автоматизированной обработки текстовой информации.

уметь:

- использовать технологию автоматизированной обработки текстовой информации для анализа языковых описаний предметных областей.
- иметь навыки: работы со специальными программными средствами автоматизированной обработки текстов.
- иметь представление: о системе языковедческих дисциплин; предмете, методах и моделях прикладной лингвистики; предмете компьютерной лингвистики.
- знать приемы и методы использования компьютерных технологий в различных видах и формах учебной и научной деятельности;

владеть

- методикой использования компьютерных технологий в предметной области;
- обладать навыками разработки педагогических технологий, создания инновационных
- программных и учебных продуктов по лингвистике, основанных на применении компьютера;

I. 4. Объем дисциплины и виды учебной работы для 3-курса

Виды учебной работы	Всего		Семестры	
	Часов	кредиты	5	6
Аудиторные занятия (всего)	**180**	**6**	**3**	**3**
В том числе				
лекции	45		23	23
Практические занятия	45		22	22
Самостоятельная работа	90		45	45
Итоговый контроль			**Экзамен**	**Экзамен**
Общая трудоемкость часов	**180**	6		

I.5. Формы контроля: зачет

Текущий контроль осуществляется лектором и преподавателем, ведущим практические занятия.

Контроль за освоением дисциплины осуществляется в каждом дисциплинарном разделе отдельно.

I.6. Технологическая карта дисциплины

Наименование дисциплины	Уровень/ступень образования (бакалавриат/магистратура)	Количество зачетных единиц/кредитов	Форма отчетности	Курс, семестр
Автоматическая обработка естественного языка	бакалавриат	6	Экзамен	3 курс 5,6семестр

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 выполнения домашних и проектных работ	Внеаудиторная или аудиторная	30	40

Выполнение заданий самостоятельную работу(Написание реферата, практических работ по задан. теме)	Выполнение определенного задания по СРС	внеаудиторная	5	15
Знание базовых понятий и примеров по основным темам и умение решать практические задачи	Устный опрос, 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**
Презентация, практикум	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: 2 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 УЧЕБНОЙ ДИСЦИПЛИНЫ

II.1. Содержания разделов дисциплины

Тема 1. Автоматическая обработка текста, системы понимания текста. Автоматическая обработка письменного текста.

Автоматизированные системы обработки текста и их архитектура. Теория подъязыков и автоматическая обработка текста.

Тема 2. Автоматическая обработка текста. Формальные языки представления текста. Формальные языки представления текста как инструмент его автоматической обработки

Тема 3. Виды автоматической обработки текстов и области ее применения. Введенные в систему тексты и/или словари размещаются в базах данных и снабжаются словоуказателями и другими индексами, позволяющими по слову или его характеристикам находить его контексты или словарные статьи, в которых оно описано. Результатом автоматической обработки текста в автоматизированных лексикографических системах являются частотные словари, конкордансы (словоуказатели с контекстами), автоматические многоязычные словари, размещаемые в базах данных используемые программами лексикографических систем в качестве справочного материала при обработке новых данных.

Тема 4. Проблема универсальных алгоритмов в автоматической обработке текста и искусственном интеллекте. Моделирование языковых явлений. Искусственный интеллект. Алгоритмы АОЕЯ текстов

Тема 5. Автоматический анализ текста. Автоматизация синтеза текста. Автоматические диалоговые системы. Лингвистическое обеспечение систем различного назначения. Автоматический синтаксический анализ

Тема 6. Проблемы использования естественного языка в системах автоматической обработки. Системы автоматического понимания текстов (АПТ) выделяются из множества систем, в которых просто используется автоматическая обработка текста, включающая техническое сжатие текста, сортировку слов по частоте, длине и т. п., любой стати-

тический анализ, исправление грамматических ошибок и другие частные задачи,

Тема 7. Анализ текста в парадигмах автоматического понимания текста. системы машинного перевода текстов с одних естественных языков на другие, системы автоматизированного поиска информации в текстах, системы автоматического анализа и синтеза устной речи

Тема 8. Практическое применение систем обработки естественно-языковых текстов. В связи с разработкой лингвистического обеспечения информационных систем разных типов в последние годы особую актуальность приобрела проблема унификации и стандартизации терминологии. Это связано с резко возросшим количеством научных и технических терминов, обслуживающих различные сферы знания. Естественно, что при этом не остается в стороне и лексикография, для которой вопросы, связанные со стандартизацией терминологии, являются в высшей степени животрепещущими. Постепенно из традиционной лексикографии все определеннее выделяется специальная сфера лексикографической деятельности, связанная с широким применением ЭВМ, – автоматическая лексикография.

Примерный план практических заданий

1. Автоматический анализ текста
2. Автоматическая обработка больших массивов документов
3. Составление глоссария
4. Морфологический анализ текста
5. Автоматический перевод текста
6. Работа с тезаурусом

II.2. Вопросы и задания для самостоятельной работы студентов

- Теория подъязыков и автоматическая обработка текста.
- Автоматические диалоговые системы.
- Динамическое программирование, скрытые марковские цепи, нейроподобные сети
- Метод чтения динамических спектрограмм как основа построения базы знаний (экспертная система) для лингвистического декодирования речевых сигналов.
- Алгоритмы транскрибирования и кодировки текстов и аллофонная база данных.
- Формантный синтез, КЛП-синтез.

III. ОЦЕНОЧНЫЕ СРЕДСТВА ПО ИТОГАМ ОСВОЕНИЯ ДИСЦИПЛИНЫ

III.1. Примерные вопросы к экзамену

1. Автоматизированная обработкаестественного языка
2. Системы автоматизированной обработки текста.
3. Какова архитектура одной из систем автоматизированной обработки текста?
4. Какие формальные языки представления текста вы знаете?
5. Каким образом реализуется автоматизация синтеза текста?
6. Каким образом реализуется автоматизация обработки текста в текстовом процессоре?
7. Какие есть способы и инструменты автоматической обработки текста в текстовом процессоре?
8. Что такое тезаурус? Для чего он применяется?
9. Как используются словари в текстовом процессоре?
10. Автоматические диалоговые системы
11. Каким образом реализуется диалог человека с автоматизированной системой?
12. Разъясните структуру любой программы тестирования.
13. Как происходит автоматическая обработка звучащей речи?
14. Каковы основные методы автоматического синтеза речи? Дайте характеристику одному из них.
15. Лингвистический подход к автоматическому распознаванию речи
16. Искусственный интеллект
17. Автоматическая обработка текста

IV. ОБРАЗОВАТЕЛЬНЫЕ ТЕХНОЛОГИИ

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

1. Андрющенко В. М. Автоматическая обработка текста // Лингвистический энциклопедический словарь – М.: Советская энциклопедия, 1990
2. Апресян Ю. Д. Избранные труды, том I. Лексическая семантика. Синонимические средства языка – М.: Языки русской культуры, 1995 217
3. Апресян Ю. Д. Экспериментальное исследование семантики русского глагола – М.: Наука, 1967
4. Апресян Ю. Д., Богуславский И. М., Иомдин Л. Л. и др. Лингвистическое обеспечение в системе автоматического перевода третьего по-

коления – М.: Научный Совет по комплексной проблеме "Кибернетика" при Президиуме АН СССР, 1978
5. Арутюнова Н. Д. Понятие пропозиции в логике и в лингвистике // Изв. АН СССР. Сер. лит. и яз. № 1, Т. 35 – М.: Наука, 1976
6. Арутюнова Н. Д., Падучева Е. В. Истоки, проблемы и категории прагматики // Новое в зарубежной лингвистике. Вып. 16 – М., 1985
7. Балли Ш. Общая лингвистика и вопросы французского языка – М.: Изд-во иностранной литературы, 1955
8. Беляева Л. Н., Откупщикова М. И. Автоматический (машинный) перевод Прикладное языкознание: Учебник – СПб, 1996- 210

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

Предусматривается следующие образовательные технологии:

- традиционные и интерактивные лекции с дискурсивной практикой обучения;
- практические и / или устные домашние задания, подготовка доклада, творческого проекта;
- самостоятельная работа студентов, в которую входит освоение теоретического материала, подготовка к практическим занятиям с использованием интернета и электронных библиотек, выполнение практических работ.

IV.3. Материально-техническое обеспечение дисциплины

Необходимо наличие компьютерного класса со следующим программным обеспечением:

- операционная система WindowsXP или Windows 7;
- пакет Microsoft Office 2007 или выше;
- Разработанные лабораторные работы, включающие в себя обучающие тексты, набор пошаговых инструкций, учебных задач и заданий, демонстрационный материал и тестовые задания, размещенные на магнитных носителях с программным обеспечением.