

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ  
Федеральное государственное автономное учреждение  
высшего профессионального образования  
"Казанский (Приволжский) федеральный университет"  
Высшая школа татаристики и тюркологии им.Габдуллы Тукая



**УТВЕРЖДАЮ**

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

"\_\_" \_\_\_\_\_ 20\_\_ г.

**Программа дисциплины**

Компьютерная верстка и автоматическая обработка текста Б1.В.ОД.7

Направление подготовки: 45.04.01 - Филология

Профиль подготовки: Литературная критика и редактирование

Квалификация выпускника: магистр

Форма обучения: очное

Язык обучения: русский

**Автор(ы):**

Хакимов Б.Э.

**Рецензент(ы):**

Хусаинова А.Х.

**СОГЛАСОВАНО:**

Заведующий(ая) кафедрой: Салехова Л. Л.

Протокол заседания кафедры No \_\_\_\_ от "\_\_\_\_" \_\_\_\_\_ 201\_\_ г

Учебно-методическая комиссия Института филологии и межкультурной коммуникации (Высшая школа татаристики и тюркологии им.Габдуллы Тукая):

Протокол заседания УМК No \_\_\_\_ от "\_\_\_\_" \_\_\_\_\_ 201\_\_ г

Регистрационный No

Казань  
2017

## 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. Хахимов Б.Э. Кафедра образовательных технологий и информационных систем в филологии Высшая школа русского языка и межкультурной коммуникации, Boulat.Hakimov@kpfu.ru

### 1. Цели освоения дисциплины

изучение основных принципов и методов автоматической обработки текстов на естественном языке (ЕЯ).

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

Данная учебная дисциплина включена в раздел " Б1.В.ОД.7 Дисциплины (модули)" основной образовательной программы 45.04.01 Филология и относится к обязательным дисциплинам. Осваивается на 2 курсе, 3 семестр.

Курс 'Компьютерная верстка и автоматическая обработка текста' является вводным в проблематику компьютерной лингвистики и построения программных систем для обработки текстов на естественном языке. Изучаются основные принципы и методы автоматической обработки текста (АОТ), а также связанные с ними модели представления знаний, применяемые в области искусственного интеллекта. Обзорно рассматривается широкий круг приложений в области АОТ, более подробно изучается задача классификации текстов на ЕЯ и задача извлечения информации из текста. Предусматривается общее ознакомление с современными системами АОТ.

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

В результате освоения дисциплины формируются следующие компетенции:

Шифр компетенции	Расшифровка приобретаемой компетенции
ОК-1 (общекультурные компетенции)	владение культурой мышления; способность к восприятию, анализу, обобщению информации, постановке цели и выбору путей ее достижения

В результате освоения дисциплины студент:

1. должен знать:

- основные уровни обработки текста на ЕЯ и существующие модели статистического, морфологического и синтаксического анализа текстов
- типичные программные системы (приложения) в области АОТ и их архитектурные особенности

2. должен уметь:

- работать с автоматическими словарями (осуществлять запросы и извлекать информацию);
- охарактеризовать принципы морфологического, синтаксического анализа (на примере одной из систем);
- осуществлять реферирование текста с использованием компьютерных средств.

3. должен владеть:

- инструментами автоматической обработки естественного языка;
- ориентированием в современных программных средствах обработки естественного языка.

4. должен продемонстрировать способность и готовность:

системного представления о методах автоматической обработки естественно-языкового текста на всех

уровнях лингвистического анализа

#### 4. Структура и содержание дисциплины/ модуля

Общая трудоемкость дисциплины составляет 2 зачетных(ые) единиц(ы) 72 часа(ов).

Форма промежуточного контроля дисциплины зачет в 3 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. Автоматическая обработка устной речи	3	1	2	0	0	
2.	Тема 2. Автоматы, формальные грамматики и языки	3	2	0	2	0	
3.	Тема 3. Ввод, распознавание и графематический анализ текста	3	3	0	2	0	
4.	Тема 4. Морфологический анализ в системах автоматической обработки текста	3	4-5	2	2	0	
5.	Тема 5. Синтаксический анализ в системах автоматической обработки текста	3	6	0	2	0	
6.	Тема 6. Семантический анализ в системах автоматического анализа текста	3	7	0	2	0	
7.	Тема 7. Словарная поддержка. Типы словарей. Компьютерные (электронные) словари.	3	8	2	0	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	
8.	Тема 8. Синтез текстов на естественном языке	3	9	0	2	0	
9.	Тема 9. Морфологическая разметка корпусов текстов. Корпус русского языка	3	10	0	2	0	
10.	Тема 10. Автоматическая обработка данных в корпусах татарского и тюркских языков	3	11	0	2	0	
	Тема . Итоговая форма контроля	3		0	0	0	зачет
	Итого			6	16	0	

#### 4.2 Содержание дисциплины

##### Тема 1. Автоматическая обработка устной речи

**лекционное занятие (2 часа(ов)):**

Автоматическая обработка устной речи

##### Тема 2. Автоматы, формальные грамматики и языки

**практическое занятие (2 часа(ов)):**

Автоматы, формальные грамматики и языки

##### Тема 3. Ввод, распознавание и графематический анализ текста

**практическое занятие (2 часа(ов)):**

Ввод, распознавание и графематический анализ текста

##### Тема 4. Морфологический анализ в системах автоматической обработки текста

**лекционное занятие (2 часа(ов)):**

Морфологический анализ в системах автоматической обработки текста

**практическое занятие (2 часа(ов)):**

Морфологический анализ в системах автоматической обработки текста

##### Тема 5. Синтаксический анализ в системах автоматической обработки текста

**практическое занятие (2 часа(ов)):**

Синтаксический анализ в системах автоматической обработки текста

##### Тема 6. Семантический анализ в системах автоматического анализа текста

**практическое занятие (2 часа(ов)):**

Семантический анализ в системах автоматического анализа текста

##### Тема 7. Словарная поддержка. Типы словарей. Компьютерные (электронные) словари.

**лекционное занятие (2 часа(ов)):**

Словарная поддержка. Типы словарей. Компьютерные (электронные) словари.

##### Тема 8. Синтез текстов на естественном языке

**практическое занятие (2 часа(ов)):**

Синтез текстов на естественном языке

##### Тема 9. Морфологическая разметка корпусов текстов. Корпус русского языка

**практическое занятие (2 часа(ов)):**

Морфологическая разметка корпусов текстов. Корпус русского языка

**Тема 10. Автоматическая обработка данных в корпусах татарского и тюркских языков**

**практическое занятие (2 часа(ов)):**

Автоматическая обработка данных в корпусах татарского и тюркских языков

**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
1.	Тема 1. Автоматическая обработка устной речи	3	1	Работа на компьютере	4	Тест
2.	Тема 2. Автоматы, формальные грамматики и языки	3	2	Работа на компьютере	4	Тест
3.	Тема 3. Ввод, распознавание и графематический анализ текста	3	3	Работа на компьютере	4	Тест
4.	Тема 4. Морфологический анализ в системах автоматической обработки текста	3	4-5	Работа на компьютере	6	Тест
5.	Тема 5. Синтаксический анализ в системах автоматической обработки текста	3	6	Работа на компьютере	6	Тест
6.	Тема 6. Семантический анализ в системах автоматического анализа текста	3	7	Работа на компьютере	4	Тест
7.	Тема 7. Словарная поддержка. Типы словарей. Компьютерные (электронные) словари.	3	8	Работа на компьютере	6	Тест
8.	Тема 8. Синтез текстов на естественном языке	3	9	Работа на компьютере	4	Тест
9.	Тема 9. Морфологическая разметка корпусов текстов. Корпус русского языка	3	10	Работа на компьютере	6	Тест

N	Раздел Дисциплины	Семестр	Неделя семестра	Виды самостоятельной работы студентов	Трудоемкость (в часах)	Формы контроля самостоятельной работы
10.	Тема 10. Автоматическая обработка данных в корпусах татарского и тюркских языков	3	11	Работа на компьютере	6	Тест
	Итого				50	

## 5. Образовательные технологии, включая интерактивные формы обучения

Знания, умения и навыки достигаются за счет использования в процессе обучения интерактивных методов компетенций у студентов:

1. лекций с применением мультимедийных технологий;
2. использование деловых игр на практических занятиях;
3. вовлечение студентов в проектную деятельность.

(Указываются образовательные технологии, используемые при реализации различных видов учебной работы.

В соответствии с требованиями ФГОС ВПО по направлению подготовки реализация компетентностног подхода должна предусматривать широкое использование в учебном процессе активных и интерактивных форм проведения занятий(компьютерных симуляций, деловых и ролевых игр, разбор конкретных ситуаций, психологические и иные тренинги) в сочетании с внеаудиторной работой с целью формирования и развития профессиональных навыков обучающихся. В рамках учебных курсов должны быть предусмотрены встречи с представителями российских и зарубежных компаний, государственных и общественных организаций, мастер-классы экспертов и специалистов.

Удельный вес занятий, проводимых в интерактивных формах, определяется главной целью (миссией) программы, особенностью контингента обучающихся и содержанием конкретных дисциплин, и в целом в учебном процессе они должны составлять не менее \_\_\_% аудиторных занятий (определяется требованиями ФГОС с учетом специфики ООП). Занятия лекционного типа для соответствующих групп студентов не могут составлять более \_\_\_% аудиторных занятий (определяется соответствующими ФГОС).

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

### Тема 1. Автоматическая обработка устной речи

Тест , примерные вопросы:

### Тема 2. Автоматы, формальные грамматики и языки

Тест , примерные вопросы:

### Тема 3. Ввод, распознавание и графематический анализ текста

Тест , примерные вопросы:

### Тема 4. Морфологический анализ в системах автоматической обработки текста

Тест , примерные вопросы:

### Тема 5. Синтаксический анализ в системах автоматической обработки текста

Тест , примерные вопросы:

### Тема 6. Семантический анализ в системах автоматического анализа текста

Тест , примерные вопросы:

## **Тема 7. Словарная поддержка. Типы словарей. Компьютерные (электронные) словари.**

Тест , примерные вопросы:

## **Тема 8. Синтез текстов на естественном языке**

Тест , примерные вопросы:

## **Тема 9. Морфологическая разметка корпусов текстов. Корпус русского языка**

Тест , примерные вопросы:

## **Тема 10. Автоматическая обработка данных в корпусах татарского и тюркских языков**

Тест , примерные вопросы:

## **Тема . Итоговая форма контроля**

Примерные вопросы к зачету:

Приложение.

### **7.1. Основная литература:**

Грудева, Е. В. Корпусная лингвистика [Электронный ресурс] : учеб. пособие / Е. В. Грудева. - 2-е изд., стер. - М. : ФЛИНТА, 2012. - 165 с. <http://znanium.com/catalog.php?bookinfo=455049>

Потапова Р.К. Новые информационные технологии и лингвистика : учебное пособие для студентов вузов, обучающихся по специальности 021800 'Теоретическая и прикладная лингвистика' направления 620200 'Лингвистика и новые информационные технологии' / Р.К. Потапова ; Моск. гос. лингвист. ун-т .? Изд. 5-е .? Москва : URSS : [ЛИБРОКОМ, 2012] .? 364 с.

Федотова Е.Л., Федотов А.А. Информационные технологии в науке и образовании : учебное пособие / Е. Л. Федотова, А. А. Федотов .? Москва : ФОРУМ : ИНФРА-М, 2011 .? 334 с.

Щипицина, Л. Ю. Информационные технологии в лингвистике [Электронный ресурс] : учеб. пособие / Л. Ю. Щипицина. ? М. : ФЛИНТА, 2013. ? 128 с. <http://znanium.com/catalog.php?bookinfo=462989>

### **7.2. Дополнительная литература:**

Галиуллин К.Р. Компьютерная лингвография / К.Р.Галиуллин. - Казань: КГУ, 1995. - 119 с.

Гатауллин Р.М. Информатика: 2 кис. 1 нче кисәк / Р.М.Гатауллин. - Казан: Мәгариф, 2006. - 191 б.

Информатика һәм мәгълүмат технологияләре терминнары: Инглизчә-татарча-русча аңлатмалы сүзлек. - Казан: Мәгариф, 2006. - 383 б.

Зубов А.В. Информационные технологии в лингвистике: учебное пособие для студентов вузов / А.В.Зубов. - М., 2004. - 205 с.

Когнитивная и компьютерная лингвистика / Ред.: Р.Г.Бухараев, В.Д.Соловьев, Д.Ш.Сулейманов. - Казань: КГУ, 1994. - 112 с.

Налимов В.В. Вероятностная модель языка. :О соотношении естественных и искусственных языков. Томск., М., 2003.

Ризванова Л.М. Квантитативная характеристика татарского слова / Л.М.Ризванова. - Казань: ТаРИХ, 2000. - 156 с.

Сабитова И.И. Словообразовательная характеристика лексики татарского языка: Компьютерно-лингвографический аспект / И.И.Сабитова. - Казань: Фикер, 2002. - 144 с.

Сөләйманов Ж.Ш., Галимжанов Ә.Ф. Информатика буенча татарча-русча-инглизчә сүзнамә / Ж.Ш.Сөләйманов, Ә.Ф.Галимжанов. - Казан: 'Сөләт' фонды, 2001. - 95 б.

Сулейманов Д.Ш., Гатиатуллин А.Р. Структурно-функциональная компьютерная модель татарских морфем / Д.Ш.Сулейманов, А.Р.Гатиатуллин. - Казань: Фэн, 2003. - 220 с.

Сулейманов Д.Ш., Хадиев Р.М., Якушев Р.С. Компьютерные информационные технологии / Д.Ш.Сулейманов, Р.М.Хадиев, Р.С.Якушев. - Казань: КГУ, 2002. - 163 с.

Татарский язык и новые информационные технологии / Ред.: Р.Г.Бухараев, В.Д.Соловьев, Д.Ш.Сулейманов. - Казань: КГУ, 1995. - 123 с.

Хроленко А.Т., Денисов А.В. Современные информационные технологии для гуманитария: практическое руководство. - М.: Флинта: Наука, 2008. - 128 с.

### **7.3. Интернет-ресурсы:**

Автоматическая обработка естественного языка - [www.aot.ru](http://www.aot.ru)

Автоматическая обработка естественного языка -  
<http://www.ssla.ru/departments/kafinf/disc/avtoobr.pdf>

АВТОМАТИЧЕСКАЯ ОБРАБОТКА ТЕКСТОВ НА ЕСТЕСТВЕННОМ ЯЗЫКЕ -

<http://clschool.miem.edu.ru/uploads/swfupload/files/011a69a6f0c3a9c6291d6d375f12aa27e349cb67.pdf>

Электронный корпус русского языка - [www.ruscorpora.ru](http://www.ruscorpora.ru)

Электронный корпус татарского языка -

[http://web-corpora.net/TatarCorpus/search/?interface\\_language=ru](http://web-corpora.net/TatarCorpus/search/?interface_language=ru)

### **8. Материально-техническое обеспечение дисциплины(модуля)**

Освоение дисциплины "Компьютерная верстка и автоматическая обработка текста" 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.

Учебно-методическая литература для данной дисциплины имеется в наличии в электронно-библиотечной системе Издательства "Лань", доступ к которой предоставлен студентам. 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.

Компьютерный класс с установленным пакетом, принтером, проектором, доской.

(Указывается материально-техническое обеспечение данной дисциплины (модуля).

Программа составлена в соответствии с требованиями ФГОС ВПО и учебным планом по направлению 45.04.01 "Филология" и магистерской программе Литературная критика и редактирование .

Автор(ы):

Хакимов Б.Э. \_\_\_\_\_

"\_\_" \_\_\_\_\_ 201\_\_ г.

Рецензент(ы):

Хусаинова А.Х. \_\_\_\_\_

"\_\_" \_\_\_\_\_ 201\_\_ г.