

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.

*подписано электронно-цифровой подписью*

### Программа дисциплины

Теория и практика машинного перевода БЗ.В.10

Направление подготовки: 032700.62 - Филология

Профиль подготовки: Прикладная филология (Татарский язык и литература, информационные технологии)

Квалификация выпускника: бакалавр

Форма обучения: очное

Язык обучения: русский

**Автор(ы):**

Хакимов Б.Э.

**Рецензент(ы):**

-

### **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 902316418

Казань

2018

## 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. Хакимов 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

Основная цель курса - познакомить студентов с основными методами автоматической обработки естественного языка (NLP - Natural Language Processing) с использованием современных информационных технологий. Данная дисциплина призвана объяснить основные причины и условия применения методов автоматического анализа языка в научных лингвистических исследованиях, прикладной, в том числе переводческой деятельности лингвиста; познакомить студентов с существующими программными продуктами для осуществления профессиональной деятельности.

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

Данная учебная дисциплина включена в раздел " Б3.В.10 Профессиональный" основной образовательной программы 032700.62 Филология и относится к вариативной части. Осваивается на 4 курсе, 7 семестр.

Данная учебная дисциплина логически и содержательно связана с другими курсами основной образовательной программы по направлению, относящимися к информационным технологиям.

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

В результате освоения дисциплины формируются следующие компетенции:

Шифр компетенции	Расшифровка приобретаемой компетенции
ОК-10 (общекультурные компетенции)	способность понимать сущность и значение информации в развитии современного информационного общества, сознавать опасности и угрозы, возникающие в этом процессе, соблюдать основные требования информационной безопасности, в том числе защиты государственной тайны
ОК-11 (общекультурные компетенции)	владение основными методами, способами и средствами получения, хранения, переработки информации, навыки работы с компьютером как средством управления информацией
ОК-12 (общекультурные компетенции)	способность работать с информацией в глобальных компьютерных сетях
ПК-13 (профессиональные компетенции)	владение базовыми навыками доработки и обработки (корректурa, редактирование, комментирование, реферирование и т. п.) различных типов текстов
ПК-2 (профессиональные компетенции)	владение базовыми навыками сбора и анализа языковых и литературных фактов с использованием традиционных методов и современных информационных технологий

В результате освоения дисциплины студент:

1. должен знать:

теоретические основы применения методов автоматического анализа языка в профессиональной, прикладной, научно-исследовательской и образовательной деятельности лингвиста;

- основные понятия и термины, относящиеся к сфере информатизации общества, науки и образования;
- основные математико-статистические методы обработки лингвистической информации;
- принципы работы специализированных программных продуктов, в том числе созданных для решения переводческих задач.

2. должен уметь:

- использовать компьютерные технические средства и стандартное программное обеспечение в профессиональных, исследовательских и образовательных целях;
- работать с основными типами профессиональных, прикладных, научно-исследовательских и учебных компьютерных программ.

3. должен владеть:

- использовать компьютерные технические средства и стандартное программное обеспечение в профессиональных, исследовательских и образовательных целях;
- работать с основными типами профессиональных, прикладных, научно-исследовательских и учебных компьютерных программ.

4. должен демонстрировать способность и готовность:

работать с компьютером как средством получения, обработки и управления информацией

#### 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. Компьютерные технологии в филологии. Введение в методы автоматического анализа языка. Ресурсы						

автоматической обработки текстов естественного языка.

7

1-3

8

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	
2.	Тема 2. Компьютерная лексикография. Формализация структуры словаря. Работа с лексикографической базой данных.	7	4-6	10	4	0	
3.	Тема 3. Корпусная лингвистика. Работа с электронными корпусами языков.	7	7-9	10	6	0	
	Тема . Итоговая форма контроля	7		0	0	0	Экзамен
	Итого			28	14	0	

#### 4.2 Содержание дисциплины

**Тема 1. Компьютерные технологии в филологии. Введение в методы автоматического анализа языка. Ресурсы автоматической обработки текстов естественного языка.**

**лекционное занятие (8 часа(ов)):**

**практическое занятие (4 часа(ов)):**

Введение. Филологические направления, в которых активно задействуются компьютерные технологии. Экскурс в проблемы автоматической обработки текста, необходимой для работы программ, анализирующих и преобразующих текстовые данные.

**Тема 2. Компьютерная лексикография. Формализация структуры словаря. Работа с лексикографической базой данных.**

**лекционное занятие (10 часа(ов)):**

Пользовательская работа с объектами базы в лексикографической практике.

**практическое занятие (4 часа(ов)):**

Общие сведения. Формализация структуры словаря. Устройство базы данных словаря. Типы информации в словаре и базе данных (БД). Объекты БД: таблицы и формы, фильтры, запросы, отчеты, макропрограммы. Пользовательская работа с объектами базы в лексикографической практике.

**Тема 3. Корпусная лингвистика. Работа с электронными корпусами языков.**

**лекционное занятие (10 часа(ов)):**

Специфика разметки языковых данных.

**практическое занятие (6 часа(ов)):**

Корпусная лингвистика (КЛ). Общие соображения. Понятия КЛ. Требования к корпусу. Специфика разметки языковых данных. Проблемы снятия неоднозначностей в корпусах текстов. Достижения КЛ. Современные проекты. Корпуса текстов on-line. Проблемы современной корпусной лингвистики.

#### 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. Компьютерные технологии в филологии. Введение в методы автоматического анализа языка. Ресурсы автоматической обработки текстов естественного языка.	7	1-3	Анализ систем и технологий в области автоматического анализа языка.	10	Домашнее задание. Реферат.
2.	Тема 2. Компьютерная лексикография. Формализация структуры словаря. Работа с лексикографической базой данных.	7	4-6	Анализ структуры татарско-русского общелексического словаря. Выделение основных зон, оформление стат	10	Домашнее задание.
3.	Тема 3. Корпусная лингвистика. Работа с электронными корпусами языков.	7	7-9	Анализ значений определенной лексики с использованием национального корпуса русского языка.	10	Домашнее задание. Реферат.
	Итого				30	

#### 5. Образовательные технологии, включая интерактивные формы обучения

Освоение курса "Методы автоматического анализа языка" предполагает использование как традиционных, так и инновационных образовательных технологий. Традиционные образовательные технологии подразумевают использование в учебном процессе таких методов работ, как лекция, семинар, практическое занятие и др. Инновационные образовательные технологии обуславливают внедрение в учебный процесс таких методов и приемов, как различные формы тренингов, деловые игры, дискуссия, моделирование ситуаций и др.

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

**Тема 1. Компьютерные технологии в филологии. Введение в методы автоматического анализа языка. Ресурсы автоматической обработки текстов естественного языка.**

Домашнее задание. Реферат. , примерные вопросы:

Анализ систем и технологий в области автоматического анализа языка. Назначение систем. Указание источника ресурсов.

### **Тема 2. Компьютерная лексикография. Формализация структуры словаря. Работа с лексикографической базой данных.**

Домашнее задание. , примерные вопросы:

Анализ структуры татарско-русского общелексического словаря. Выделение основных зон, оформление статьи с использованием языка разметки. Результат - формальное представление статьи словаря.

### **Тема 3. Корпусная лингвистика. Работа с электронными корпусами языков.**

Домашнее задание. Реферат. , примерные вопросы:

Анализ значений определенной лексики с использованием национального корпуса русского языка. Исследование всех значений лексики в определенном периоде времени. Построение графика.

### **Тема . Итоговая форма контроля**

Примерные вопросы к экзамену:

Примерные вопросы к зачету /экзамену:

- 1) Лингвистические компьютерные технологии.
- 2) История компьютерной лингвистики.
- 3) Компьютерные методы лингвистических исследований.
- 4) Автоматический анализ текста.
- 5) Лингвистические модели.
- 6) Формализация языковой структуры.
- 7) Компьютерная лексикография, электронные словари.
- 8) Квантитативная лингвистика. Частотные словари.
- 9) Корпусная лингвистика.
- 10) Использование корпусов текстов в научных исследованиях.
- 11) Лингвистические ресурсы и поиск в Интернет.
- 12) Лингвистические технологии информационного поиска.
- 13) Системы машинного перевода.

#### **7.1. Основная литература:**

Федотова Е.Л., Федотов А.А. Информационные технологии в науке и образовании: Учебное пособие / Е.Л. Федотова, А.А. Федотов. - М.: ИД ФОРУМ: НИЦ ИНФРА-М, 2015. - 336 с. <http://znanium.com/catalog.php?bookinfo=487293>

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.

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>

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>



## 7.2. Дополнительная литература:

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

Хроленко А.Т., Денисов А.В. Современные информационные технологии для гуманитария: практическое руководство. - М.: Флинта: Наука, 2008. - 128 с.  
<http://znanium.com/catalog.php?bookinfo=320764>

## 7.3. Интернет-ресурсы:

Британский национальный корпус - <http://www.natcorp.ox.ac.uk>

Компания 'Аби' - <http://abbyu.ru>

Лаборатории общей и компьютерной лексикографии МГУ - <http://lexigraph.nm.ru/library.htm>

Национальный корпус русского языка - <http://www.ruscorpora.ru>

НИИ 'Прикладная семиотика' АН РТ - <http://ips.antat.ru>

Филологический факультет МГУ - <http://www.philol.msu.ru/~lex/main.htm>

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

Освоение дисциплины "Теория и практика машинного перевода" предполагает использование следующего материально-технического обеспечения:

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.

Освоение дисциплины "Методы автоматического анализа языка" предполагает использование следующего материально-технического обеспечения:

Мультимедийная аудитория, состоящая 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, audi, 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.

Программа составлена в соответствии с требованиями ФГОС ВПО и учебным планом по направлению 032700.62 "Филология" и профилю подготовки Прикладная филология (Татарский язык и литература, информационные технологии) .

Автор(ы):

Хакимов Б.Э. \_\_\_\_\_

"\_\_" \_\_\_\_\_ 201\_\_ г.

Рецензент(ы):

"\_\_" \_\_\_\_\_ 201\_\_ г.