




МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ
РОССИЙСКОЙ ФЕДЕРАЦИИ
федеральное государственное бюджетное образовательное
учреждение высшего образования
«ИРКУТСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ»
ФГБОУ ВО «ИГУ»

Факультет иностранных языков
Кафедра английской филологии

Утверждаю
Декан факультета иностранных языков


Кузнецова О.В.
«30» апреля 2020 г.

Рабочая программа дисциплины

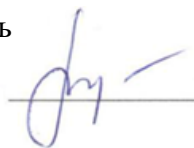
Б1.В.ДВ.9.2 Корпусная лингвистика

Направление подготовки	45.03.02 «Лингвистика»
Профиль	«Теория и методика преподавания иностранных языков и культур (первый иностранный язык – китайский, второй иностранный язык – английский)»
Квалификация (степень) выпускника	бакалавр
Форма обучения	очная

Согласовано с УМК ИФИЯМ

Протокол № 10 от «30» апреля 2020 г.

Председатель



О.Л. Михалева

Рекомендовано кафедрой:

Протокол № 6 от «10» февраля 2020 г.

Зав.кафедрой востоковедения и регионоведения
АТР



Кремнёв Е.В.

Содержание

	стр.
1. Цели и задачи дисциплины	3
2. Место дисциплины в структуре ОПОП	3
3. Требования к результатам освоения дисциплины	4
4. Объем дисциплины и виды учебной работы	5
5. Содержание дисциплины	6
5.1 Содержание разделов и тем дисциплины	6
5.2 Разделы дисциплины и междисциплинарные связи с обеспечиваемыми (последующими) дисциплинами	7
5.3 Разделы и темы дисциплины и виды занятий	8
6. Перечень семинарских, практических занятий и лабораторных работ	9
6.1. План самостоятельной работы студентов	10
6.2. Методические указания по организации самостоятельной работы студентов	11
7. Примерная тематика курсовых работ	13
8. Учебно-методическое и информационное обеспечение дисциплины:	14
а) основная литература;	14
б) дополнительная литература;	14
в) программное обеспечение;	14
г) базы данных, поисково-справочные и информационные системы	16
9. Материально-техническое обеспечение дисциплины	16
10. Образовательные технологии	16
11. Оценочные средства (ОС)	18

1. Цели и задачи дисциплины

Целью дисциплины Б1.В.ДВ.9.2 «Корпусная лингвистика» является формирование углубленных знаний об образовательном потенциале Интернет и компьютерных технологий, навыков обработки русскоязычных и иноязычных текстов в производственно-практических целях, разработки средств информационной поддержки лингвистических областей знаний, а также ознакомление с основами корпусной лингвистики и электронными иноязычными корпусами.

Освоение дисциплины должно обеспечить возможность формирования следующих компетенций: ОПК-11, ОПК-13, ОПК-14, ОПК-16, ОПК-20, ПК-25.

Задачи дисциплины

- 1) формирование углубленных навыков работы с информационными ресурсами: поиск, сбор, анализ, оценка, синтез и интеграция нового знания в существующую систему знаний обучаемого;
- 2) оказание методической помощи в изучении и практическом использовании современных компьютерных и телекоммуникационных технологий для профессиональной работы.

Дисциплина служит также основой для развития творческого потенциала студентов.

Задачи дисциплины заключаются в подготовке студента к выполнению следующих профессиональных задач:

- изучение и критический анализ в профессиональных целях материалов исследований в области прикладной лингвистики;
- совершенствование компетенций в области научной организации труда и интегрированных решений профессионально-значимых прикладных задач;
- проведение самостоятельных исследований в рамках прикладной лингвистики в целях повышения своего общепрофессионального уровня, использования результатов исследований в практической деятельности и подготовки к продолжению образования;
- проведение постоянной информационно-поисковой работы с целью совершенствования компетенции специалиста в области приложения лингвистической теории к процессу решения прикладных задач.

2. Место дисциплины в структуре ОПОП

Дисциплина Б1.В.ДВ.9.2 «Корпусная лингвистика» изучается на 2 курсе в 4 семестре и относится к Б1.В.ДВ вариативному блоку образовательной программы по направлению подготовки 45.03.02 "Лингвистика", профиль «Теория и методика преподавания иностранных языков и культур» дисциплин по выбору.

Изучение дисциплины предполагает базовые знания, приобретенные обучаемыми, в ходе изучения курсов информатики, математики и профессионально ориентированного перевода (первый иностранный язык).

Входные знания, умения и компетенции студента, необходимые для изучения дисциплины:

- обладать базовыми навыками публичного выступления на общие темы;
- владеть базовыми навыками применения современных программных средств обработки и редактирования информации, в том числе на иностранных языках;

- быть способным самостоятельно организовать свою профессиональную деятельность, оценивать ее результаты;
- быть способным анализировать взаимосвязи явлений и фактов действительности на базовом уровне исследований;
- обладать знаниями, умениями и навыками по основным лингвистическим теоретическим дисциплинам первого иностранного языка.

3. Требования к результатам освоения дисциплины

Процесс изучения дисциплины направлен на формирование следующих компетенций:

- владеть навыками работы с компьютером как средством получения, обработки и управления информацией (ОПК-11);
- быть способным работать с электронными словарями и другими электронными ресурсами для решения лингвистических задач (ОПК-13);
- владеть основами современной информационной и библиографической культуры (ОПК-14);
- владеть стандартными методиками поиска, анализа и обработки материала исследования (ОПК-16);
- быть способным решать стандартные задачи профессиональной деятельности на основе информационной и библиографической культуры с применением информационно-лингвистических технологий и с учетом основных требований информационной безопасности (ОПК-20);
- иметь навыки работы с компьютером как средством получения, обработки и управления информацией (ПК-25).

В результате изучения дисциплины студент должен

Знать

- специфику работы и основные приемы работы с компьютером, с различными носителями информации, базами данных, компьютерными сетями, электронными словарями и другими электронными ресурсами;
- основы современных методов научного исследования и основы информационной и библиографической культуры.

Уметь

- соотносить лингвистические данные с технологическими процессами их обработки и применения;
- выбирать и адекватно применять информационно-поисковые и экспертные системы, системы представления знаний, синтаксического и морфологического анализа, автоматического синтеза и распознавания речи, обработки лексикографической информации и автоматизированного перевода, автоматизированными системами идентификации и верификации личности;
- решать стандартные задачи профессиональной деятельности на основе информационной и библиографической культуры.

Владеть

- основными навыками работы с компьютером для получения, обработки и управления информацией приемами самостоятельного изучения языковых явлений в теоретическом и прикладном аспектах;
- навыками работы с различными носителями информации, базами данных, компьютерными сетями, с электронными словарями и другими электронными ресурсами;
- навыками решения стандартных задач профессиональной деятельности на основе информационной и библиографической культуры с применением информационно-

лингвистических технологий и с учетом основных требований информационной безопасности;

- современными методами научного исследования, информационной и библиографической культурой.

4. Объем дисциплины и виды учебной работы

Вид учебной работы	Всего часов / зачетных единиц	Семестр
		4
Аудиторные занятия (всего)	42	42
В том числе:		
Лекции	20	20
Практические занятия (ПЗ)	20	20
Семинары (С)		
Лабораторные работы (ЛР)		
Контроль самостоятельной работы (КСР)	2	2
Самостоятельная работа (всего)	30	30
В том числе:		
Курсовой проект (работа)		
Расчетно-графические работы		
Реферат (при наличии)		
<i>Другие виды самостоятельной работы</i>		
Работа над учебным материалом, включая выполнение домашних заданий		
Вид промежуточной аттестации		ЗаО
Контактная работа (всего)		
Общая трудоемкость: часы	72	72
зачетные единицы	2	2

5. Содержание дисциплины

5.1. Содержание разделов и тем дисциплины

№ раздела	Наименование раздела	Наименование темы
1.	Введение в корпусную лингвистику	Тема 1. Основные понятия корпусной лингвистики. Понятие корпуса. Корпус и текст. Корпусная лингвистика как дисциплина. Виды и свойства корпусов. Web as a corpus. Применение корпусов. Значение корпусов. Корпусы и компьютерная лингвистика. Разметка. Конкорданс, конкордансер. Критика корпусной лингвистики.

2.	История создания и типология англоязычных корпусов	Создание англоязычных корпусов – периодизация. Первые машиночитаемые корпусы. Брауновский корпус. Ланкастерско-Осло-Бергенский корпус. Синтаксически размеченные корпусы. The Penn Treebank. Британский национальный корпус. Международный корпус английского языка. Корпус «Банк английского». Корпус современного американского английского. Оксфордский корпус английского. Корпусы TenTen. Корпус Google Books Ngram. Корпусы с семантической разметкой. FrameNet. Groningen Meaning Bank.
3	Создание корпусов	<p>Проектирование корпусов. Техпроцесс создания корпусов. Машинная обработка текстов: токенизация, лемматизация, стемминг, парсинг. Автоматическая разметка. Проблема авторских прав. Стандартизация. Техники бутстрэпнга.</p> <p>Корпусные технологии. Ретроспективный анализ корпусных технологий. Корпусные технологии. Количественный и качественный контент-анализ корпуса текстов. Корпусный инструментарий. Разметка. Виды разметок. Инструментальные средства разработки и анализа корпуса текстов и медиа. Типы корпусов. Устные и письменные, одноязычные и многоязычные, аннотированные и неаннотированные корпусы.</p>
4.	Корпусы русского языка	Упсальский корпус русского языка. Тюбингенские корпусы русских текстов. Компьютерный корпус текстов русских газет конца XX века. Система Диктум-1. Большой Корпус русского языка. Машинный фонд русского языка. Корпус русского литературного языка. ХАНКО. НКРЯ. OpenCorpora.

5.2. Разделы дисциплины и междисциплинарные связи с обеспечиваемыми (последующими) дисциплинами

№№ п/п	Наименование обеспечиваемых (последующих) дисциплин	№ № разделов данной дисциплины, необходимых для изучения обеспечиваемых (последующих) дисциплин
1.	Б1.В.ДВ.2.2 Применение современных информационных коммуникационных технологий в научно-исследовательской деятельности	1,2,3,4,5,6,7,8
2.	Б1.В.ДВ.5.1 Практикум по научно-исследовательской	1,2,3,4,5,6,7,8

	деятельности	
3.	Б1.В.ДВ.2.1 Современные образовательные технологии	1,2,3,4,5,6,7,8
4.	Б1.В.ДВ.5.2 Практикум по профессиональной педагогической коммуникации	1,2,3,4,5,6,7,8
5.	Б1.В.ДВ.3.2 Профессионально ориентированный перевод (первый иностранный язык) (английский)	1,2,3,4,5,6,7,8
6.	Б1.В.ДВ.4.2 Профессионально ориентированный перевод (второй иностранный язык) (немецкий)	1,2,3,4,5,6,7,8

5.3. Разделы и темы дисциплин и виды занятий

№ п/п	Наименование раздела	Наименование темы	Виды занятий в часах			
			Лекции	Практические занятия	СРС	Всего
1.	Введение в корпусную лингвистику История создания и типология англоязычных корпусов	Тема 1.	2	4	4	10
		Тема 2.	2	2	4	8
2.	Создание корпусов	Тема 3.	2	2	3	7
		Тема 4..	2	2	3	7
3.	Корпусы русского языка Введение в корпусную лингвистику История создания и типология англоязычных корпусов	Тема 5.	2	2	3	7
		Тема 6.	2	2	3	7
		Тема 7.	2	2	3	7
4.	Создание корпусов	Тема 8.	2	2	3	7
		Тема 9.	4	2	4	10
Всего часов:			20	20	30	70

6. Перечень практических занятий

№ п/п	№ раздела и темы дисциплины	Наименование семинаров, практических и лабораторных работ	Трудовое время (часы)	Оценочные средства	Формируемые компетенции
1	2	3	4	5	6
Раздел 1. Понятие информации, информационных технологий и информационных систем.					
1.	Тема 1. Информация, её формы, свойства и классификация.	Практические занятия	4	БРС	ОК-11, ОПК-11, ОПК-13, ОПК-14, ОПК-16, ОПК-20, ПК-25.
2.	Тема 2. Методы решения задач использованием информационных технологий.	Практические занятия	2	БРС	ОК-11, ОПК-11, ОПК-13, ОПК-14, ОПК-16, ОПК-20, ПК-25.
Раздел 1. Понятие информации, информационных технологий и информационных систем.			6	Защита презентации и беседа по теме	ОК-11, ОПК-11, ОПК-13, ОПК-14, ОПК-16, ОПК-20, ПК-25.
Раздел 2. Компьютерная лингвистика.					
3.	Тема 3. Информационные технологии в лингвистике. Машинный перевод.	Практические занятия	2	БРС	ОК-11, ОПК-11, ОПК-13, ОПК-14, ОПК-16, ОПК-20, ПК-25.
4.	Тема 4. Системы автоматического понимания текста и извлечения информации.	Практические занятия	2	БРС	ОК-11, ОПК-11, ОПК-13, ОПК-14, ОПК-16, ОПК-20, ПК-25.
Раздел 2. Компьютерная лингвистика.			4	Практические задания	ОК-11, ОПК-11, ОПК-13, ОПК-14, ОПК-16, ОПК-20, ПК-25.
Раздел 3. Актуальные направления прикладной лингвистики.					
5.	Тема 6. Компьютерная лексикография.	Практические занятия	2	БРС	ОК-11, ОПК-11, ОПК-13, ОПК-14, ОПК-16, ОПК-20, ПК-25.
6.	Тема 7. Корпусные технологии.	Практические занятия	2	БРС	ОК-11, ОПК-11, ОПК-13, ОПК-14, ОПК-16, ОПК-20, ПК-25.
7.	Тема 8. Компьютерные технологии в обучении.	Практические занятия	2	БРС	ОК-11, ОПК-11, ОПК-13, ОПК-14, ОПК-16, ОПК-20, ПК-25.
Раздел 3. Актуальные направления прикладной лингвистики.			6	Практические задания. Контрольный срез по полученным знаниям	ОК-11, ОПК-11, ОПК-13, ОПК-14, ОПК-16, ОПК-20, ПК-25.
Раздел 4. Основы компьютерных телекоммуникаций.					
8.	Тема 9. Компьютерные сети.	Практические занятия	2	БРС	ОК-11, ОПК-11, ОПК-13, ОПК-14, ОПК-16, ОПК-20, ПК-25.

9.	Тема 10. Принципы работы в современной информационной среде.	Практические занятия	2	БРС	ОК-11, ОПК-11, ОПК-13, ОПК-14, ОПК-16, ОПК-20, ПК-25.
Раздел 4. Основы компьютерных телекоммуникаций			4	Контрольный срез по полученным знаниям	ОК-11, ОПК-11, ОПК-13, ОПК-14, ОПК-16, ОПК-20, ПК-25.

6.1. План самостоятельной работы студентов

№ недель	Тема	Вид самостоятельной работы	Задание	Рекомендуемая литература	Количество часов
Раздел 1. Понятие информации, информационных технологий и информационных систем.					
1.	Тема 1. Информация, ее формы, свойства и	Работа над учебным материалом, включая выполнение домашних заданий	Работа литературой и источниками	См. Раздел «Учебно-методическое и информационное обеспечение дисциплины»	2
2.	Тема 2. Методы решения задач с использованием информационных технологий.	Работа над учебным материалом, включая выполнение домашних заданий	Работа литературой и источниками	См. Раздел «Учебно-методическое и информационное обеспечение дисциплины»	2
Раздел 2. Компьютерная лингвистика.					
3.	Тема 3. Информационные технологии в лингвистике. Машинный перевод.	Работа над учебным материалом, включая выполнение домашних заданий	Работа литературой и источниками	См. Раздел «Учебно-методическое и информационное обеспечение дисциплины»	2
4.	Тема 4. Системы автоматического понимания текста и извлечения информации.	Работа над учебным материалом, включая выполнение домашних заданий	Работа литературой и источниками	См. Раздел «Учебно-методическое и информационное обеспечение дисциплины»	2
Раздел 3. Актуальные направления прикладной лингвистики.					
5.	Тема 6. Компьютерная лексикография.	Работа над учебным материалом, включая выполнение домашних заданий	Работа литературой и источниками	См. Раздел «Учебно-методическое и информационное обеспечение дисциплины»	2
6.	Тема 7. Корпусные технологии.	Работа над учебным материалом, включая выполнение домашних заданий	Работа литературой и источниками	См. Раздел «Учебно-методическое и информационное обеспечение дисциплины»	2
7.	Тема 8. Компьютерные технологии в обучении.	Работа над учебным материалом, включая выполнение домашних заданий	Работа литературой и источниками	См. Раздел «Учебно-методическое и информационное обеспечение дисциплины»	2
Раздел 4. Основы компьютерных телекоммуникаций.					
8.	Тема 8. Компьютерные сети.	Работа над учебным материалом, включая выполнение	Работа литературой и источниками	См. Раздел «Учебно-методическое и информационное	2

		домашних заданий		обеспечение дисциплины»	
9.	Тема Принципы работы современной информационной среде.	9. Работа над учебным материалом, включая выполнение домашних заданий	Работа с литературой и источниками	См. Раздел «Учебно-методическое и информационное обеспечение дисциплины»	4

6.2. Методические указания по организации самостоятельной работы студентов

Методические рекомендации и учебно-методические материалы по СРС

представляют собой комплекс материалов и подходов, позволяющих студенту оптимальным образом организовать процесс изучения данной дисциплины в рамках СРС, а также рекомендации преподавателям по различным аспектам организации СРС.

Особое место в освоении дисциплины «Информационные технологии в лингвистике» занимает самостоятельная работа студентов. Самостоятельная работа – это вид учебной деятельности, одна из организационных форм обучения, протекающая вне непосредственного контакта с преподавателем (дома, в лаборатории ТСО) или управляемая преподавателем опосредованно через предназначенные для этой цели учебные материалы.

Внеаудиторная самостоятельная работа предполагает не только работу с текущим лекционным материалом и первоисточниками, но и индивидуальную учебно-исследовательскую работу, нацеленную на формирование навыков самостоятельного изучения исследуемой лингвистической проблематики. В связи с чем выполнение заданий в ходе самостоятельной работы предусматривает разноаспектный формат, который включает в себя работу в научной библиотеке, с источниками (в том числе обработка данных авторефератов кандидатских диссертаций, научных лингвистических журналов), а также работу в сети Интернет. Последний вид работы предполагает применение навыков критического мышления для отбора материала для дальнейшей презентации на семинарских занятиях в виде разнообразных речевых (устных и письменных) продуктов.

Контроль качества выполненных заданий предусматривается в специально отведенные и особым образом организованные аудиторные часы.

Самостоятельная работа студентов предполагает следующие виды деятельности:

- изучение учебной, научной и справочной литературы по темам, представленным в разделах курса и реферирование базовых положений по основным темам;
- практическое освоение компьютерных программ, применяемых для автоматизированной обработки текстовых массивов и выполнение практических заданий;
- реализация учебных проектов (командных, групповых и индивидуальных);
- подготовка устных и письменных речевых продуктов для выступления на семинарских занятиях по исследуемому вопросу.

Таким образом, в процессе самостоятельной работы студента встречаются все типы заданий по дидактической цели: познавательные, практические, обобщающие.

По уровню проблемности самостоятельная работа студента предполагает выполнение репродуктивных, репродуктивно-исследовательских, исследовательских и продуктивных (творческих) заданий.

По методам научного познания выполняемые задания для самостоятельной работы студента по дисциплине «Современные информационные коммуникационные технологии в лингвистике» являются теоретическими, экспериментальными и проектно-творческими,

направленными на моделирование, наблюдение, классификацию, обобщение, систематизацию полученных знаний в учебных продуктивных проектах.

По типам решаемых задач. Здесь возможны разные классификации, например, типы задач: познавательные, творческие, исследовательские.

В процессе подготовки к лекции студенту необходимо в ходе самостоятельной работы

- тщательно проработать тему предыдущей лекции для ответа на вопросы преподавателя, заданные с целью повторения пройденного материала;
- подготовить сообщения на основе рекомендуемой дополнительной литературы и с привлечением Интернет-ресурсов;
- составить глоссарий лингвистических терминов;
- самостоятельно подобрать примеры из разнообразных валидных источников, иллюстрирующие теоретические положения в рамках предметной тематики.

В процессе подготовки к практическому занятию студенту необходимо в ходе самостоятельной работы

- изучить соответствующие лекции и разделы рекомендуемых учебников и дополнительной литературы, в том числе первоисточников;
- апробировать и освоить рекомендованные программой курса компьютерные программы или подобрать предлагаемые в других источниках альтернативные программы;
- подготовить устные и письменные речевые продукты для презентации перед аудиторией.

При подготовке к контрольным занятиям студенту необходимо в ходе самостоятельной работы

- тщательно проработать и усвоить материал лекций и разделы рекомендованных учебников и научной литературы по соответствующим темам;
- обратить особое внимание на определения основных научных понятий, формулировки проблем и примерам решения практических задач, приводимых на лекциях и семинарах;
- решить в ходе самостоятельной подготовки к контрольной работе ряда аналогичных заданий из рекомендованных сборников.

7. Примерная тематика исследовательских проектов

1. корпусная лингвистика
2. корпус
3. общий (общезыковой) корпус
4. специализированный корпус
5. параллельный корпус
6. метаданные
7. аннотации
8. тэг
9. репрезентативность корпуса
10. "Googleology"
11. конкорданс
12. тип
13. токен
14. лексема
15. лемма
16. словоформа
17. ключевое слово
18. коллокация
19. коллигация
20. парсинг
21. время создания Брауновского корпуса
22. количество слов в Брауновском корпусе
23. состав Брауновского корпуса
24. основной недостаток Брауновского корпуса
25. суть закона Ципфа применительно к частотности слов
26. что общего между более чем половиной слов в Брауновском корпусе
27. какой корпус стал моделью для Ланкастерско-Осло-Бергенского корпуса
28. сколько компьютерных корпусов было разработано к 1990 г. (приблизительно)
29. treebank
30. NP (синтаксис)
31. VP (синтаксис)
32. время создания Penn Treebank
33. количество слов в Penn Treebank
34. состав Penn Treebank
35. как кодируется иерархическая структура в Penn Treebank
36. количество слов в British National Corpus
37. соотношение устной и письменной речи в BNC
38. время создания BNC
39. разметка в BNC
40. тэггер
41. корпусный менеджер / менеджер корпусов
42. Xaira
43. цель проекта International Corpus of English
44. количество слов в каждом подкорпусе ICE
45. соотношение устной и письменной речи в ICE
46. виды разметки в ICE
47. количество слов в COCA
48. количество жанров в COCA

49. wildcard
50. n-грамма
51. синхронический корпус
52. диахронический корпус
53. национальный корпус
54. мониторинговый (мониторинговый, динамический) корпус
55. сбалансированный корпус
56. Sketch Engine
57. как собираются большие корпуса (>2 млрд. слов)
58. недостатки корпусов, сформированных из "очищенных" веб-страниц
59. как происходит автоматическое создание специализированных корпусов
60. чем отличается термин от ключевого слова
61. какие тексты содержатся в Google Books Ngram Corpus
62. вид представления данных в GBNC
63. виды разметки в GBNC
64. абсолютная частота
65. относительная частота
66. к какому виду корпусов относятся FrameNet и GMB
67. на какой теории основан проект FrameNet
68. семантический фрейм
69. "глубина" разметки FrameNet
70. количество семантических фреймов во FrameNet
71. количество аннотированных примеров во FrameNet
72. назовите не менее 4 видов разметки в Groningen Meaning Bank
73. количество токенов в последнем релизе GMB
74. бутстрэппинг - основная идея
75. game with a purpose
76. примеры экстралингв. разметки (не менее 3)
77. главное отличие стемминга от лемматизации
78. BootCaT
79. на основе чего BootCaT отбирает тексты
80. что использует BootCaT для получения web-страниц
81. Уппсальский корпус - объем
82. Уппсальский корпус - тексты
83. есть ли treebank на русском языке?
84. Машинный фонд русского языка - тексты
85. акцентуированный корпус
86. Хельсинкский аннотированный корпус русского языка - объем
87. Хельсинкский аннотированный корпус русского языка - виды разметки
88. НКРЯ - объем основного корпуса
89. НКРЯ - назовите не менее 4 подкорпусов
90. НКРЯ - хронологический диапазон текстов
91. OpenCorpora - объем
92. OpenCorpora - главное преимущество
93. как переводчики используют паралл. корпуса
94. как комп. лингвисты используют паралл. корпуса
95. выравнивание / стыковка
96. параллельный корпус в НКРЯ - структура
97. LF aligner
98. проект Татоэба

99. российск. проект по созд. устных корпусов на русском и др. языках
100. корпус устной речи в НКРЯ - состав
101. СОСА: walk как глагол
102. СОСА: все словоформы walk
103. СОСА: синонимы к walk
104. СОСА: все слова, начинающиеся на walk (walker, walking, ...)
105. СОСА: биграммы walk + любое слово
106. СОСА: биграммы walk + любое слово, кроме away
107. СОСА: walk или stroll
108. СОСА: walk в конце повествовательного предложения
109. AntConc
110. AntConc: что нужно для корректной работы с текстовыми файлами
111. AntConc: для чего кнопка Clone Results
112. AntConc: как экспортировать результаты во внешний файл
113. AntConc: Concordance Plot (для чего)
114. AntConc: File View (для чего)
115. AntConc: Clusters (для чего)
116. AntConc: N-Grams (для чего)
117. AntConc: Collocates (для чего)
118. AntConc: Word List (для чего)
119. AntConc: Keyword List (для чего)
120. откуда AntConc "знает", какие слова являются ключевыми
121. AntConc: "негативные" ключевые слова
122. может ли AntConc определять леммы слов
123. возможности TreeTagger для текстов на русском языке
124. рекомендуемый текстовый редактор
125. Python: открыть текстовый файл (без чтения)
126. Python: прочитать открытый текстовый файл
127. Python: сохранить строку в текстовый файл
128. Python: разбить текст на отдельные строки
129. Python: разбить текст на "слова"
130. Python: "собрать" текст из отдельных строк
131. Python: "собрать" текст из отдельных слов
132. Python: структура данных для слов/строк в определенном порядке
133. Python: структура данных для набора уникальных словоформ
134. Python: структура данных для лемм и соотв. словоформ

135. 8. Учебно-методическое и информационное обеспечение дисциплины

а) основная

1. Баранова Е.В. Информационные технологии в образовании [Электронный ресурс] / Е.В. Баранова. – Москва : Лань, 2016. – Режим доступа: http://e.lanbook.com/books/element.php?pl1_id=81571 .

б) дополнительная

1. Полат Е.С. Новые педагогические и информационные технологии в системе образования: Учеб. пособие для вузов. [Текст] / ред. Е.С. Полат. – М. : Академия, 2002. – 272 с. (23 экземпляра)
2. Полат Е.С. Педагогические технологии дистанционного обучения: учеб. пособие для вузов [Текст] / ред. Е.С. Полат. – М. : Академия, 2006. – 400 с. (76 экземпляров)
3. Кудинов Ю.И. Основы современной информатики: учеб. [Текст] / Ю.И. Кудинов, Ф.Ф. Пашенко. – Москва: Лань, 2016. – 256 с. – Режим доступа: http://e.lanbook.com/books/element.php?pl1_id=86016 .
4. Кудинов Ю.И. Практикум по основам современной информатики [Электронный ресурс] / Ю.И. Кудинов, Ф.Ф. Пашенко, А.Ю. Келина. – Москва: Лань, 2011. – 350 с. – Режим доступа: http://e.lanbook.com/books/element.php?pl1_id=68471.
5. Роберт И.В. Теория и методика информатизации образования (психолого-педагогический и технологический аспекты). [Электронный ресурс] – Электрон. дан. – М. : Издательство "Лаборатория знаний", 2014. – 398 с. – Режим доступа: <http://e.lanbook.com/book/66196>

Согласовано НБ ИГУ

в) программное обеспечение

1. «**Kaspersky Endpoint Security** для бизнеса - Стандартный Russian Edition. 1500-2499 Node 2 year Educational Renewal License». – Срок действия: от 2017.12.13 до 2020-01-21.
2. **Microsoft Desktop Education** Allng License / software Assurance Pack Academic OLV 1 License Level E Enterprise. 1 Year до 2018-11-30 (сейчас). Следующая подписка с 2018-11-30 по 2019-11-30.
3. **Drupal 7.5.4**. Условия правообладателя (Лицензия GPL-2.0 - ware free). Условия использования по ссылке: https://www.drupal.org/project/terms_of_use. Обеспечивает работу портала электронного портфолио студентов и аспирантов ИГУ <http://eportfolio.isu.ru>. Срок действия: бессрочно.
4. **Moodle 3.5.1**. – Условия правообладателя (ware free). Условия использования по ссылке: <https://ru.wikipedia.org/wiki/Moodle>. Обеспечивает работу информационно-образовательной среды <http://belca.isu.ru>. Срок действия: бессрочно.
5. **Google Chrome 54.0.2840**. Браузер – Условия правообладателя (ware free). Условия использования по ссылке: https://ru.wikipedia.org/wiki/Google_Chrome/_. Срок действия: бессрочно.
6. **Mozilla Firefox 50.0**. Браузер – Условия правообладателя (ware free). Условия использования по ссылке: <https://www.mozilla.org/ru/about/legal/terms/firefox/>_. Срок действия: бессрочно.
7. **Opera 41**. Браузер – Условия правообладателя (ware free). Условия использования по ссылке: <http://www.opera.com/ru/terms>_. Срок действия: бессрочно.

8. **PDF24Creator 8.0.2.** Приложение для создания и редактирования документов в формате PDF. –Условия правообладателя (ware free). Условия использования по ссылке: https://en.pdf24.org/pdf/lizenz_en_de.pdf . Срок действия: бессрочно.
9. **VLC Player 2.2.4.** Свободный кроссплатформенный медиаплеер. Условия правообладателя (ware free). – Условия использования по ссылке: <http://www.videolan.org/legal.html>. Срок действия: бессрочно.
10. **BigBlueButton.** Открытое программное обеспечение для проведения веб-конференции. Условия правообладателя (ware free). Условия использования по ссылке: <https://ru.wikipedia.org/wiki/BigBlueButton>. Обеспечивает работу отдельного модуля Moodle 3.2.1 для работы ИОС. Срок действия: бессрочно.
11. **АСТ-Тест Plus 4.0** (на 75 одновременных подключений) и Мастер-комплект (АСТ-Maker и АСТ-Converter) (ежегодно обновляемое ПО), лицензий - 75шт.. Договор №1392 от 30.11.2016, срок действия: 3 года.
12. **Sumatra PDF.** свободная программа, предназначенная для просмотра и печати документов в форматах PDF, DjVu[4], FB2, ePub, MOBI, CHM, XPS, CBR/CBZ, для платформы Windows. Условия правообладателя (Лицензия GNU GPL 3-ware free). Условия использования по ссылке: https://ru.wikipedia.org/wiki/Sumatra_PDF. Срок действия: бессрочно.
13. **Media player home classic.** Свободный проигрыватель аудио- и видеофайлов для операционной системы Windows. Условия правообладателя (Лицензия GNU GPL - ware free). Условия использования по ссылке: https://ru.wikipedia.org/wiki/Media_Player_Classic. Срок действия: бессрочно.
14. **AIMP.** Бесплатный аудиопроигрыватель с закрытым исходным кодом, написанный на Delphi. Условия правообладателя (ware free). Условия использования по ссылке: <https://www.aimp.ru/>. Срок действия: бессрочно.
15. **Speech analyzer.** Условия правообладателя (ware free). Условия использования по ссылке: <http://www-01.sil.org/computing/sa/index.htm>. Срок действия: бессрочно.
16. **Audacity.** Программа анализа звучащей речи. Условия правообладателя (Лицензия - GNU GPL v2). Условия использования по ссылке: <https://ru.wikipedia.org/wiki/Audacity>. Срок действия правообладателя: бессрочно.
17. **Inkscape.** свободно распространяемый векторный графический редактор, удобен для создания как художественных, так и технических иллюстраций. Условия правообладателя: Лицензия -**GNU GPL 2 (ware free)**. Условия использования по ссылке: <https://inkscape.org/ru/download/>. Срок действия: бессрочно.
18. **Skype 7.30.0.** Бесплатное проприетарное программное обеспечение с закрытым кодом, обеспечивающее текстовую, голосовую и видеосвязь через Интернет между компьютерами, опционально используя технологии пиринговых сетей. Условия правообладателя (Лицензия Adware) - Условия использования по ссылке: <https://www.skype.com/ru/about/>. Срок действия: бессрочно.
19. **Adobe Acrobat Reader.** Бесплатный мировой стандарт, который используется для просмотра, печати и комментирования документов в формате PDF. Условия правообладателя:Условияиспользованияпоссылке: <https://get.adobe.com/ru/reader/otherversions/>. Срок действия: бессрочно.
20. **Circles.** Программа интерактивной визуализации многоуровневых данных: числовых значений или древовидные структуры. Условия правообладателя (Лицензия BSD). Условия использования по ссылке: <https://carrotsearch.com/circles/free-trial/>.

г) базы данных, информационно-справочные и поисковые системы

1. Электронно-библиотечная система (ЭБС).
2. Университетская информационная система РОССИЯ (УИС РОССИЯ).
3. База данных Eastview.com.
4. Научная электронная библиотека ELIBRARY.RU.
5. Электронная коллекция книг Оксфордского Российского фонда (eBookcollectionOxfordRussiaFund).

9. Материально-техническое обеспечение дисциплины

Факультет располагает современной информационно-технологической инфраструктурой, обеспечивающей проведение всех видов дисциплинарной и междисциплинарной подготовки, лабораторной, практической и научно-исследовательской работы студентов, предусмотренных учебными планами.

В институте имеются в наличии 11 компьютерных классов (135 персональных компьютера, имеющих выход в Интернет); 5 компьютерных классов и 35 мультимедийных класса, 1 конференц-зал оснащены стационарным мультимедийным оборудованием для аудиовизуальной демонстрации материалов лекционных курсов (41 проектор, 41 экран, 6 интерактивных досок); в 2-х мультимедийных классах находятся 2 телевизора с жидкокристаллическим экраном.

В учебном процессе применяется также передвижное презентационное оборудование: 34 ноутбука и 7 проекторов. Все компьютеры оснащены необходимыми комплектами лицензионного программного обеспечения.

Компьютерные классы объединены в локальную сеть; обеспечивается доступ к информационным ресурсам, к различным базам данных; в читальных залах открыт доступ к справочной и научной литературе, к периодическим изданиям. Кроме этого, учебный корпус оснащен точками свободного доступа wi-fi, которыми активно пользуются студенты и сотрудники.

10. Образовательные технологии

Использование интерактивной модели обучения предусматривает моделирование жизненных ситуаций, использование ролевых игр, совместное решение проблем. Исключается доминирование какого-либо участника учебного процесса или какой-либо идеи. Создается среда образовательного общения, которая характеризуется открытостью, взаимодействием участников, равенством их аргументов, накоплением совместного знания. Из объекта воздействия студент становится субъектом взаимодействия, он сам активно участвует в процессе обучения, следуя своим индивидуальным маршрутом. Преподаватель мотивирует обучающихся к самостоятельному поиску информации. Задачей преподавателя становится создание условий для инициативы студентов. Преподаватель отказывается от роли фильтра, пропускающего через себя учебную информацию, и выполняет функцию помощника в работе, одного из источников информации.

Учебный процесс, опирающийся на использование интерактивных методов обучения, организуется с учетом включенности в процесс познания всех студентов группы, каждый вносит свой особый индивидуальный вклад, в ходе работы идет обмен знаниями, идеями, способами деятельности, обязательна обратная связь, возможность взаимной оценки и контроля.

К методам интерактивного обучения относятся те, которые способствуют вовлечению в активный процесс получения и переработки знаний. Интерактивные методы пробуждают у обучающихся интерес, поощряют активное участие каждого в учебном процессе, обращаются к чувствам каждого обучающегося, способствуют эффективному

усвоению учебного материала, оказывают многоплановое воздействие на обучающихся, осуществляют обратную связь (ответная реакция аудитории), формируют у обучающихся мнения и отношения, формируют жизненные навыки, способствуют изменению поведения.

Удельный вес занятий, проводимых в интерактивных формах, определяется главной целью (миссией) программы, особенностью контингента обучающихся и в целом в учебном процессе составляет не менее 30% аудиторных занятий, что соответствует требованиям ФГОС.

11. Оценочные средства

11.1. Оценочные средства для входного контроля

Типовое задание

Дайте определение следующих ключевых понятий:

- информация;
- технологии;
- информационной технологии; • информатизация;
- технологии обработки лингвистической информации; • технологии обработки текстовой информации;
- технологии обработки числовой информации;
- технологии обработки графической информации; • технологии обработки звуковой информации;
- технологии работы в глобальных сетях;
- социальные информационные технологии.

11.2 Оценочные средства текущего контроля **Типовое задание для беседы по теме:**

Раздел 1. Понятие информации, информационных технологий и информационных систем.

Вопросы для самопроверки

1. Основные понятия корпусной лингвистики. Понятие корпуса. Корпус и текст.
2. Корпусная лингвистика как дисциплина. Критика корпусной лингвистики.
3. Виды и свойства корпусов. Web as a corpus.
4. Применение корпусов. Значение корпусов. Корпусы и компьютерная лингвистика.
5. Разметка и метаданные. Конкорданс, конкордансер.
6. Создание англоязычных корпусов – периодизация.
7. Первые машиночитаемые корпуса. Брауновский корпус. Ланкастерско-Осло-Бергенский корпус.
8. Синтаксически размеченные корпуса. The Penn Treebank.
9. Британский национальный корпус.
10. Международный корпус английского языка. Корпус «Банк английского».
11. Корпус современного американского английского.
12. Оксфордский корпус английского. Корпусы TenTen.
13. Корпус Google Books Ngram.
14. Корпусы с семантической разметкой. FrameNet. Groningen Meaning Bank.
15. Проектирование корпусов. Техпроцесс создания корпусов.
16. Машинная обработка текстов: токенизация, лемматизация, стемминг, парсинг.
17. Автоматическая разметка. Техники бутстрэппинга. Автоматическая семантическая разметка. USAS.
18. Проблема авторских прав. Стандартизация.
19. Упсальский корпус русского языка. Машинный фонд русского языка. Хельсинкский аннотированный корпус (ХАНКО).
20. НКРЯ.
21. OpenCorpora.
22. Параллельные корпуса – возможности и применение. Параллельный подкорпус НКРЯ. Проект Татоеба.
23. Web-интерфейсы к корпусам (СОСА, НКРЯ) – возможности, достоинства и недостатки.
24. Регулярные выражения. Обработка текста с помощью языка программирования Python (2 или 3).
25. Корпусные менеджеры. AntConc. Xaira. Sketch Engine.
26. Достоинства и недостатки корпусов, собранных автоматически из сети Интернет. BootCaT.

Демонстрационный вариант практического задания Раздел

3. Актуальные направления прикладной лингвистики.

1. Подберите тексты 2 разных жанров для анализа.
2. Воспользуйтесь видео инструкцией и создайте частотные словари на базе отобранных текстов в 2х форматах (Microsoft Word и Excel). <http://it.lang-study.com/sozdanie-chastotnogo-slovary-a-word-excel/>
3. Сопоставьте процедуры применения программ Microsoft Word и Excel.
4. Проанализируйте, полученные результаты с целью выявления лингвистически значимой

информации.

5. Сделайте выводы, где полученные результаты анализа могут найти своё применение.

Демонстрационный вариант контрольной работы

Задание: *выберите верный ответ:*

1. Какое из высказываний является определением прикладной лингвистики?
 - a) область языкознания, направленная на объективное установление состояния отдельного языка, его истории и закономерностей;
 - b) область языкознания, связанная с использованием компьютерных инструментов — программ, технологий организации и обработки данных — для моделирования функционирования языка в тех или иных условиях;
 - c) область языкознания, связанная с разработкой методов решения практических задач использования языка;
 - d) область языкознания, связанная с применением компьютерных моделей языка в лингвистике и в смежных с ней дисциплинах.

2. К направлениям компьютерной лингвистики не относится
 - a) компьютерная лексикография;
 - b) компьютерно-опосредованная коммуникация;
 - c) системы обработки естественного языка;
 - d) машинный перевод.

3. Информатика – это
 - a) наука об управлении, связи и переработке информации;
 - b) наука о накоплении, обработке и передаче информации с помощью ЭВМ;
 - c) наука о накоплении, обработке и передаче информации о строении языка с помощью ЭВМ;
 - d) наука об использовании компьютерных инструментов для моделирования функционирования языка в тех или иных условиях.

4. Разное количество информации в одном и том же сообщении для разных людей зависит не от...
- a) накопленных ими знаний;
 - b) уровня понимания сообщения;
 - c) их интереса к сообщению;
 - d) их уровня владения компьютерной техникой.
5. Следствие третьей информационной революции состоит в том, что...
- a) информация становится общедоступной;
 - b) информацию можно автоматически обрабатывать и передавать с большой скоростью;
 - c) информацию можно легко найти с помощью инструментов поиска и совместно производить;
 - d) информация может накапливаться.
6. Для современного человека преобладающей является...
- a) звуковая информация;
 - b) визуальная (символьная) информация;
 - c) вкусовая и тактильная информация;
 - d) визуальная (образная) информация.

7. Адекватность информации – это ...
- степень соответствия информации объективной реальности окружающего мира;
 - степень соответствия информации, полученной потребителем, тому, что автор вложил в ее содержание;
 - достаточность информации для принятия решения;
 - степень соответствия информации текущему моменту времени.
8. Естественный язык – это ...
- знаковая система, используемая человеком с момента рождения;
 - знаковая система, используемая человеком в непринужденной обстановке;
 - знаковая система, созданная для естественных наук;
 - знаковая система, стихийно возникшая и закрепившаяся в обществе.
9. К устройствам ввода данных не относится
- сканер;
 - принтер;
 - клавиатура;
 - цифровой фотоаппарат.
10. OCR – это ...
- система автоматического распознавания символов;
 - система переводческой памяти;
 - система машинного перевода;
 - функция текстового процессора.
11. Реферат – это ...
- связный текст, который кратко выражает тему, предмет, цель, методы и результаты исследования;
 - процесс составления содержания документа (книги, статьи, патента на изобретение и др.);
 - краткое изложение содержания документа, дающее общее представление о его теме;
 - краткий текст, выполняющий сигнальную функцию (информирует о том, что есть публикация на определенную тему).
12. Метод автоматического аннотирования, при котором важные слова выделяются в заголовке, подзаголовке, начале и конце текста, называется ...
- статистическим;
 - логико-семантическим;
 - позиционным;
 - функциональным.
13. Совокупность специально отобранных текстов, размеченных по различным лингвистическим параметрам и обеспеченных системой поиска, называется ...
- базой данных;
 - словарем;
 - информационным массивом;
 - корпусом.

14. Разметка бывает ...

- a) морфологической; синтаксической; семантической и просодической;
- b) полнотекстовой и фрагментной;
- c) синхронической и диахронической;
- d) звуковой, письменной, смешанной.

15. Требования к корпусам

- a) полнота, адекватность, актуальность, компьютерная поддержка;
- b) устойчивость, тиражируемость, адаптируемость, оптимальность временных параметров, комфорт пользователя;
- c) репрезентативность, полнота, экономичность, структуризация, компьютерная поддержка;
- d) полнота, экономичность, достоверность, структуризация, компьютерная поддержка.

16. Корпусный менеджер ...

- a) обеспечивает сортировку результатов поиска, статистические подсчеты, составление списков слов на основе корпуса;
- b) это специальная программа поиска по корпусу;
- c) это человек, составляющий корпуса и управляющий ими;
- d) это специальная программа подготовки текстов к их включению в корпус.

17. Одна из основных проблем компьютерного анализа речи состоит в том, что ...

- a) невозможно создать искусственный интеллект;
- b) компьютер не умеет работать со смыслом;
- c) у компьютера нет дополнительных источников информации (ситуация, контекст, прошлый опыт в данной области и т.п.);
- d) разработчики не желают делиться своими профессиональными секретами.

18. Электронный словарь – это ...

- a) введенный в компьютер бумажный словарь, снабженный средствами поиска и отображения информации;
- b) организованное собрание слов с комментариями, в которых описываются особенности структуры и/или функционирования этих слов;
- c) организованное собрание слов с описанием их значения, особенностей употребления, структурных свойств, сочетаемости, соотношения с лексическими системами других языков и т.д.;
- d) словарь в специальном машинном формате, предназначенный для применения на ЭВМ пользователем или компьютерной программой.

19. Что не относится к процессу и понятию машинного перевода?

- a) междисциплинарность;
- b) использование машинных средств;
- c) принципиальное сходство этапов понимания и синтеза текста;
- d) учет языковых и экстралингвистических знаний.

20. Требования к системам МП включают ...

- a) устойчивость, тиражируемость, адаптируемость, оптимальность временных параметров, комфорт пользователя;
- b) полнота, адекватность, актуальность, достоверность;
- c) репрезентативность, полнота, экономичность, адекватность, компьютерная поддержка;
- d) репрезентативность, полнота, экономичность, структуризация, компьютерная поддержка.

21. Аббревиатура CALL относится к ...

- a) науке об использовании компьютерных инструментов для моделирования функционирования языка в тех или иных условиях;
- b) обучению иностранному языку;
- c) обучению языку с помощью компьютера;
- d) использованию компьютеров в обучении.

22. Сущность когнитивно-интеллектуального подхода в компьютерном обучении состоит в том, что ...

- a) программы ориентированы на обучающегося, дают свободу выбора уровня и типа действий;
- b) программы построены по формуле стимул — реакция;
- c) обучающемуся отводится роль объекта обучения;
- d) в нем используются программы-тренажеры обучению языку с помощью компьютера.

23. К обучающим программным средствам не относятся ...

- a) тестирующие программы;
- b) энциклопедии;
- c) программы-ассемблеры;
- d) учебные игры.

24. Компьютерный учебник – это ...

- a) программа, предлагающая пользователю вопрос и несколько вариантов ответов на него;
- b) программа формирования автоматического навыка выполнения определенных коммуникативных действий путем многочисленных повторов;
- c) программы, предназначенные для представления учебного материала;
- d) программно-методический комплекс, позволяющий самостоятельно освоить учебный курс или его большой раздел.

25. Что не относится к компьютерным обучающим программам?

- a) заменяют преподавателя;
- b) организация и выполнение рутинной работы;
- c) повышение активности обучаемого;
- d) создание возможностей для самообразования.

11.3. Оценочные средства для промежуточной аттестации (в форме дифференцированного зачёта)

Текущий и промежуточный контроль осуществляется в рамках балльно-рейтинговой системы оценки учебной работы студентов (БРС). БРС представлен в прямой доступ студентам (<http://belca.isu.ru/course/view.php?id=137>) в виде электронной автоматической интерактивной таблицы, что позволяет мобильно, гибко и эффективно ранжировать студентов по результатам их учебной деятельности, а также мотивировать студентов на активную позицию по регулированию результатов баллорейтинга.

БРС включает оценку по следующим показателям:

Посещаемость занятий (1)

Активность на занятиях и в рамках СРС (до 5 баллов)

Рубежный контроль.

Добор баллов (для пропустивших занятия по уважительным причинам).

Бонусные баллы (за реализацию наиболее качественных проектов в рамках СРС – до 5 баллов).

4 семестр

Вид промежуточной аттестации – Дифференцированный зачёт.

Форма проведения

Промежуточная аттестация студентов выставляется по итогам баллорейтинга исходя их следующих критериев:

СООТВЕТСТВИЕ РЕЙТИНГОВЫХ ПРОЦЕНТОВ АКАДЕМИЧЕСКОЙ ОЦЕНКЕ

Общее количество набранных процентов	Академическая оценка
60-72	3 (удовлетворительно)
73-85	4 (хорошо)
86-100	5 (отлично)

Разработчик:



старший преподаватель

Садовникова
Оксана
Эдуардовна

Программа рассмотрена на заседании кафедры английской филологии

10 февраля 2020
года Протокол №6

Зав.кафедрой востоковедения и

регионоведения АТР



Кремнёв Е.В.

Настоящая программа не может быть воспроизведена ни в какой форме без предварительного письменного разрешения кафедры-разработчика программы

