

Министерство науки и высшего образования Российской Федерации
Федеральное государственное бюджетное образовательное учреждение
высшего образования
«Смоленский государственный университет»

Кафедра русского языка

«Утверждаю»
Проректор по учебно-методической работе
_____ Ю.А. Устименко
«15» октября 2020 г.

Рабочая программа дисциплины
Б1.В.ДВ.01.01 Автоматизированная обработка текстовых массивов

Направление подготовки: **44.03.05 Педагогическое образование (с двумя профилями подготовки)**

Направленность (профиль): **Немецкий язык, Английский язык**

Форма обучения - очная

Курс – 2

Семестр – 3

Всего зачетных единиц – 3, часов – 108

Форма отчетности: зачет – 3 семестр

Программу разработал:
к.ф.н., доцент Лунькова Е.С.

Одобрена на заседании кафедры
«08» октября 2020 г. Протокол № 2

Заведующий кафедрой _____

Смоленск
2020

1. Место дисциплины в структуре ОП

Дисциплина «Автоматизированная обработка текстовых массивов» относится к дисциплинам по выбору части, формируемой участниками образовательных отношений, учебного плана. Освоение дисциплины тесно связано с другими лингвистическими дисциплинами, поскольку предполагает использование знаний, полученных в курсах общего языкознания, русского языка, а также фонетики, лексикологии и грамматики изучаемого (английского, немецкого, французского) языка, с точки зрения функционального использования тех или иных языковых единиц, при проведении исследовательских работ студентов; при решении отдельных учебных задач с привлечением большого массива языковых данных для верификации отдельных положений лингвистической науки.

Программа по курсу «Автоматизированная обработка текстовых массивов» составлена в соответствии с требованиями государственного образовательного стандарта, предполагающего целый ряд профессиональных качеств специалистов данного направления подготовки: способность работать с информацией различного характера на русском и иностранных языках для решения профессиональных задач; способность использовать извлеченные из текстовых массивов данные для формирования научных знаний в предметных областях.

2. Планируемые результаты обучения по дисциплине

Компетенция	Индикаторы достижения
ПК-5 Способен использовать научные знания в предметной области (немецкий язык) в процессе формирования предметной компетенции обучающихся в рамках реализации основной общеобразовательной программы	Знать: основной категориальный аппарат курса: лингвистический алгоритм, язык программирования, базы данных, автоматическое аннотирование текста, конкорданс, тезаурус, ключевые слова, корпус текстов и др. Уметь: применять полученные знания при анализе текстов различной функциональной направленности, составлять пользовательские подкорпуса в пределах национальных корпусов языков и исследовательские корпуса с помощью специальных программ в соответствии с поставленными задачами в изучаемой предметной области. Владеть: компьютерными программами, предназначенными для автоматизированной обработки текстов; навыками работы в корпусах русского и иностранных языках.
ПК-3. Способен применять современные информационно-коммуникационные технологии в учебном процессе для достижения планируемых результатов обучения	Знать: открытые образовательные ресурсы и принципы разработки электронных образовательных ресурсов на доступных электронных платформах; методы поиска достоверной информации на основе Интернет технологий; принципы работы с основными текстовыми, табличными и графическими редакторами; Уметь: применять принципы и методы разработки электронных образовательных ресурсов и обеспечивать их реализацию;

	<p>использовать Интернет технологии для поиска достоверной информации в целях ее включения в образовательный процесс; использовать приемы и соблюдение правил работы со средствами ИКТ;</p> <p>Владеть: навыками разработки и реализации части учебной дисциплины в форме электронного образовательного ресурса в рамках основной общеобразовательной программы основного общего и среднего общего образования; навыками применения электронных средств сопровождения образовательного процесса; навыками создания новых документов с использованием необходимых редакторов.</p>
--	---

3. Содержание дисциплины

Компьютерная лингвистика: задачи, подходы, ресурсы.

Разделы компьютерной лингвистики.

Приложения компьютерной лингвистики. Сложности моделирования естественного языка.

Общие этапы и модули обработки текстов. Лингвистические ресурсы: построение и применение.

Корпуса языков: многообразие, принципы построения. Национальные корпуса языков. Национальный корпус русского языка. Разметка текстов в корпусе, виды разметки.

Корпуса английского языка. Корпуса других языков.

Морфологический анализ текстов.

Обзор модулей морфологического анализа. Методы хранения словарей.

Проблема морфологической омонимии: виды и способы решения. Особенности омонимии в разных языках.

Извлечение информации из текстов. Специфика задач, подходы к решению проблемы, виды извлекаемой информации. Инструментальные системы.

Обзор вероятностных тематических моделей. Основы тематического регулирования. Регуляризация. Интерпретируемость тем. Определение числа тем. Модальности. Зависимости. Иерархии тем. Тематическая сегментация.

4. Тематический план

№	Разделы и темы	Всего часов	Формы занятий		
			Лекционные занятия	Практические занятия	Самостоятельная работа
1.	Автоматизированная обработка текстов как часть компьютерной лингвистики. Разделы компьютерной лингвистики.	8	2	6	8
2.	Основные понятия корпусной лингвистики.	8	4	4	8
3.	Национальный корпус русского языка. Структура и функции.	10	2	6	10
4.	Британский национальный корпус. Корпус современного американского английского. Корпуса других языков.	10	2	4	10
5.	Обзор пользовательских корпусов для проведения исследования. Таггирование и разметка текста, специфика видов разметки.	6	4	6	10
6.	AntConc: специфика корпуса, разметка текста, функция «Regular expressions».	10	2	6	8
	Зачет. Конкурс исследовательских проектов (презентаций)	2			6
	Итого	108	16	32	60

5. Виды образовательной деятельности

Занятия лекционного типа

Лекция 1.

Компьютерная лингвистика: задачи, подходы, ресурсы.

Приложения компьютерной лингвистики. Сложности моделирования естественного языка.

Автоматизированная обработка текстов как часть компьютерной лингвистики. Разделы компьютерной лингвистики.

Лекция 2.

Основные понятия корпусной лингвистики. Корпуса языков: многообразие, принципы построения. Национальные корпуса языков. Национальный корпус русского языка. Разметка текстов в корпусе, виды разметки.

Лекция 3.

Национальный корпус русского языка. Структура и функции. Морфологический анализ текстов.

Лекция 4.

Британский национальный корпус. Корпус современного американского английского. Корпуса других языков. Таггирование и разметка текста, специфика видов разметки в корпусах разных языков.

Лекция 5.

Обзор пользовательских корпусов для разных задач исследования. Извлечение информации из текстов. Специфика задач, подходы к решению проблемы, виды извлекаемой информации. Инструментальные системы.

Лекция 6.

Общие этапы и модули обработки текстов. Лингвистические ресурсы: построение и применение. Методы хранения словарей. Проблема морфологической омонимии. Особенности омонимии в разных языках.

Практические занятия

Практическое занятие 1.

План

1. История компьютерной лингвистики, основные направления современной компьютерной лингвистики.
2. Компьютерная лексикография. Электронные словари.
3. Моделирование общения. Гипертекст и поисковые системы. Моделирование структуры сюжета.
3. Основные понятия компьютерной лингвистики. Лингвистические базы данных: виды и назначение в лингвистических исследованиях.

Практическое занятие 2.

План

1. Основные понятия корпусной лингвистики.
2. Токенизация и ее разновидности. Лемматизация, специфика лемматизации в разных языках.
3. Таггирование. Виды тагов. Зависимость тагов от структуры естественного языка.

Практическое занятие 3.

План

1. Национальный корпус русского языка (НКРЯ).
2. Структура и функции НКРЯ.
3. Коэффициент частотности (IPM).
4. Лексико-грамматический поиск.
5. Поиск точных форм.
6. Поиск по грамматическим характеристикам.
7. Биграммы.
8. Функция «Частота».
9. Функция «Графики».
10. Настройка пользовательских подкорпусов.

Практическое занятие 4.

План

1. Британский национальный корпус.
2. Корпус современного американского английского языка (СОСА). Другие корпуса английского языка.
3. Корпуса немецкого и французского языков.
4. Корпуса других языков (чешский, китайский, испанский).
5. Глобальные интернет-корпуса.

Практическое занятие 5.

План

1. Новые информационные технологии и лингвистика.
2. Обзор современных программ для создания пользовательских корпусов.
3. Специфика линейных корпусов разметки текста.
4. Таггирование и разметка текста, специфика видов разметки.

Практическое занятие 6.

План

1. Программа AntConc: специфика корпуса, функции поиска.
2. Разметка текста, N-граммы.
3. Функция «Regular expressions», формулы программирования на языке Perl для автоматизированного извлечения информации из текстового массива в корпусе.

Самостоятельная работа

Практические занятия

Практическое занятие 1.

1. Определите, к какому виду прикладных программ относятся перечисленные ниже программные продукты:

- 1) текстовые редакторы;
- 2) графические редакторы;
- 3) электронные таблицы;
- 4) веб-редакторы;
- 5) веб-браузеры.

Opera, MS Excel, MS FrontPage, Adobe Photoshop, Corel WordPerfect.

Как эти программы можно использовать в лингвистических исследованиях?

Практическое занятие 2.

1. Ответьте на вопросы.

Что такое знак? В чем различие между знаками языка и знаками речи?

В каких сферах ограничениями пословного распознавания звучащей речи можно пренебречь?

Какие артикуляционные признаки звуков вам известны?

Представители каких профессий должны быть задействованы в создании сложных систем анализа звучащей речи?

Практическое занятие 3.

1. Национальный корпус русского языка (НКРЯ). Составьте пользовательский запрос по лексико-тематической группе «Профессии» в двух подкорпусах с помощью лексико-грамматического поиска. Сравните полученные результаты.

2. Проследить актуальные процессы в изменении семантики: русский - российский, товарищ – гражданин. Изменение семантики слова «приватизация»: сравнить функционирование этого существительного с глаголом «приватизировать» и выяснить, какая часть речи деполитизируется чаще.

- 1) История метафор: валютный коридор (собрать данные, проверить анализ сочетаемости – какое выражение наиболее частотно («вести валютный коридор» и как имя ситуации, пропозициональное значение)? Когда появилось это выражение?

Морфология:

- 1) Определите количество перехода одной части речи в другую: субстантивация, адвербиализация.
- 2) Определите с помощью корпуса средства выражения грамматического значения категории одушевленности-неодушевленности (на примере слова персонаж).

Практическое занятие 4.

1. Напишите эссе (объемом до 500 слов) на одну из предложенных тем:

- История создания Корпуса современного американского языка.
- История создания Национального британского корпуса.
- Специфика Исторического корпуса американского языка.
- Критерии составления языковых корпусов.

- Параметры релевантности поиска в языковых корпусах.
- Понятие «сбалансированности» языкового корпуса.

Практическое занятие 5.

1. Выясните, какие существительные (прилагательные, глаголы, наречия и т.д.) наиболее частотны в американском английском.

2. Определите функции грамматически многозначных слов в русском языке: ТАК – в несколько этапов: выписать из Грамматического словаря Зализняка части речи, функции которых может выполнять ТАК (наречие, часть союза и др.); подобрать по несколько примеров на каждый случай.

3. Ответьте на вопросы.

Что такое формальная и смысловая релевантность поиска? Как различие этих понятий отражается на результатах поиска?

Охарактеризуйте два типа информационно-поисковых систем: документальные и фактографические.

В чем состоят различия поисковых систем с ручным и автоматическим индексированием? Приведите примеры систем обоих типов.

Практическое занятие 6:

1. Работа с программой AntConc. Составление пользовательского подкорпуса на русском языке. Организация поисковых запросов студентов с учетом их научно-исследовательских интересов (работа по курсовым проектам).

2. Охарактеризуйте структуру словарной статьи, использующейся для автоматизированного извлечения информации из текстовых массивов.

Опишите зону морфологических сведений. Какие кодировки используются для обозначения частей речи и представления морфологической информации?

Чем различаются зона семантических и зона лексических сведений компьютерной словарной статьи? Проиллюстрируйте различия примерами.

Посетите сайт www.merriam-webster.com. Введите слово culture в строку поиска.

Определите зоны словарной статьи для этого словаря и представьте результаты вашего анализа в виде таблицы.

Зоны словарной статьи	Данные для слова culture в электронном словаре Merriam Webster

Посетите сайт www.ozhegov.org. Введите слово культура в строку поиска. Определите зоны словарной статьи для этого словаря и представьте результаты вашего анализа в виде таблицы.

Зоны словарной статьи	Данные для слова культура в электронном словаре Ожегова

Сравните количество зон словарной статьи в двух рассмотренных словарях: в каком словаре их больше? Какие нужные, на ваш взгляд, зоны отсутствуют в словарях? С каким словарем вам было удобнее работать и почему?

6. Критерии оценивания результатов освоения дисциплины

6.1. Оценочные средства и критерии оценивания для текущей аттестации

1. Практические занятия

Текущий контроль над освоением студентом дисциплины «Автоматизированная обработка текстовых массивов» проводится на практических занятиях.

Критерии оценивания участия работы на лабораторном занятии

Студент демонстрирует уверенное знание учебного материала по тематике лабораторного занятия, усвоил основную литературу, знаком с дополнительной литературой, рекомендованной программой, свободно оперирует приобретенными знаниями по автоматизированной обработке текстовых массивов.	Зачтено
Студент демонстрирует низкий уровень знаний учебного материала по тематике лабораторного занятия, слабо усвоил основную литературу, не знаком с дополнительной литературой, рекомендованной программой, не может оперировать приобретенными знаниями по автоматизированной обработке текстовых массивов.	Не зачтено

Тестирование

Образцы тестовых заданий для текущего контроля знаний

Тест

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.
- 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.
- Для современного человека преобладающей является...
- a) звуковая информация;
 - b) визуальная (символьная) информация;
 - c) вкусовая и тактильная информация;
 - d) визуальная (образная) информация.
- Адекватность информации — это ...

- a) степень соответствия информации объективной реальности окружающего мира;
- b) степень соответствия информации, полученной потребителем, тому, что автор вложил в ее содержание;
- c) достаточность информации для принятия решения;
- d) степень соответствия информации текущему моменту времени.

Машинный синтаксис — это ...

- a) правила строения имен;
- b) правила построения слов в более сложные структуры;
- c) соотношение слова и его значения;
- d) правила перевода письменного символа в устный.

Естественный язык — это ...

- a) знаковая система, используемая человеком с момента рождения;
- b) знаковая система, используемая человеком в непринужденной обстановке;
- c) знаковая система, созданная для естественных наук;
- d) знаковая система, стихийно возникшая и закрепившаяся в обществе.

10. Волапюк — это...

- a) специализированный язык науки;
- b) родной язык одного из малочисленных племен;
- c) неспециализированный искусственный язык;
- d) система символического кодирования.

11. Какие из следующих приложений не являются текстовыми редакторами?

- a) MS Excel;
- b) Corel WordPerfect;
- c) MS Works;
- d) Adobe InCopy.

12. Microsoft Word не включает...

- a) функции настольных издательских систем;
- b) функцию удалённого доступа;
- c) функцию редактирования графических объектов;
- d) шаблоны типовых таблиц.

13. К устройствам ввода данных не относится

- a) сканер;
- b) принтер;
- c) клавиатура;
- d) цифровой фотоаппарат.

14. OCR — это ...

- a) система автоматического распознавания символов;
- b) система переводческой памяти;
- c) система машинного перевода;
- d) функция текстового процессора.

15. Реферат — это...

- a) связный текст, который кратко выражает тему, предмет, цель, методы и результаты исследования;
- b) процесс составления содержания документа (книги, статьи, патента на изобретение и др.);
- c) краткое изложение содержания документа, дающее общее представление о его теме;
- d) краткий текст, выполняющий сигнальную функцию (информирует о том, что есть публикация на определенную тему).

16. Слово, относящееся к основному содержанию текста и повторяющееся в

нем несколько раз, в автоматическом реферировании называется ...

- a) лейтмотивом;
- b) термином;
- c) символом;
- d) ключевым словом.

17. Метод автоматического аннотирования, при котором важные слова выделяются в заголовке, подзаголовке, начале и конце текста, называется ...

- a) статистическим;
- b) логико-семантическим;
- c) позиционным;
- d) функциональным.

18. Совокупность специально отобранных текстов, размеченных по различным лингвистическим параметрам и обеспеченных системой поиска, называется

...

- a) базой данных;
- b) словарем;
- c) информационным массивом;
- d) корпусом.

19. Разметка бывает ...

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

20. УНК — это ...

- a) корпус естественного языка, представительный по отношению ко всему языку;
- b) универсальный национальный код;
- c) собрание текстов, которое существует в Интернете;
- d) собрание текстов, размеченных по различным лингвистическим параметрам и обеспеченных системой поиска.

21. 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.

22. 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.

23. ПОД — это ...

- a) вид информационно-поисковой системы;
- b) специальная программа поиска по корпусу;
- c) поисковый образ документа;
- d) поисковая оценка данных.

24. 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.

25. 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.

26. К зонам словарной статьи не относится

- a) лексический вход (вокабула, лемма);
- b) зона грамматической информации;
- c) зона стилистических помет;
- d) словник.

27. Что включает в себя понятие АСПОТ?

- a) словарь в специальном машинном формате, предназначенный для применения на ЭВМ пользователем;
- b) компьютерные версии хорошо известных словарей (Вебстер, Коллинз, Ожегов...);
- c) словарь в специальном машинном формате, предназначенный для применения на ЭВМ компьютерной программой;
- d) словари, предназначенные для обычного пользователя.

28. Что не относится к понятию термина?

- a) слово (словосочетание) метаязыка науки, а также областей конкретной практической деятельности человека;
- b) понятие задается через свойства, реализуемые в системе;
- c) использование основывается не на интуиции, а на четких определениях;
- d) сопоставляется, как правило, несколько значений.

29. 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.

30. Типовая парадигма лексемы в автоматическом морфологическом анализе — это ...

- a) последовательность букв от начала словоформы, общая для всех словоформ;
- b) элементы, описывающие формоизменение конкретной лексемы,
- c) совокупность наборов машинных окончаний;
- d) совпадение основ разных слов.

31. 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.

32. 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.

Критерии оценивания тестирования

Для получения зачета по тесту студент должен правильно выполнить на не менее чем 50% тестовых заданий. В случае, если количество правильно выполненных тестовых заданий составляет меньше 50%, студент не получает зачета.

6.2. Оценочные средства и критерии оценивания для промежуточной аттестации

По итогам работы на лабораторных занятиях на последнем из них студенту выставляется зачет. Средствами оценивания являются задания для самостоятельной работы на практических занятиях.

Критерии оценивания:

1. Зачет выставляется по результатам работы студента в течение семестра согласно Положению о текущем контроле успеваемости и промежуточной аттестации студентов в федеральном государственном бюджетном образовательном учреждении высшего образования «Смоленский государственный университет» (утверждено приказом ректора от 24 апреля 2014 г. №01-36).

2. Для получения зачета студент должен:

- пропустить не более 49% занятий;

- изучить рекомендованную рабочей программой основную и дополнительную литературу;

- выполнить все определенные программой задания для самостоятельной работы;

- написать эссе;

- активно участвовать во всех формах аудиторных занятий.

Оценка «**зачтено**» предполагает:

Хорошее знание основных терминов и понятий курса;

Хорошее знание и владение методами и средствами решения задач;

Последовательное изложение материала курса;

Умение формулировать некоторые обобщения по теме вопросов;

Достаточно полные ответы на вопросы;

Умение использовать фундаментальные понятия при ответе.

Оценка «**не зачтено**» предполагает:

Неудовлетворительное знание основных терминов и понятий курса;

Отсутствие логики и последовательности в изложении материала курса;

Неумение формулировать отдельные выводы и обобщения по теме вопросов;

Неумение использовать фундаментальные понятия при ответах.

7. Перечень основной и дополнительной учебной литературы

7.1. Основная литература

1. Информатика и лингвистика: Учебное пособие/Волосатова Т.М., Чичварин Н.В. - М.: НИЦ ИНФРА-М, 2016. - 196 с.: 60x90 1/16. - (Высшее образование: Бакалавриат) (Переплёт 7БЦ) ISBN 978-5-16-010977-0

<http://znanium.com/bookread2.php?book=508097>

7.2. Дополнительная литература

1. Аверина С.А., Азарова И.В., Алексеева Е.Л. и др. Прикладное языкознание. – СПб., 1996.
2. Баранов А.Н. Введение в прикладную лингвистику. – М., 2013.
3. Вежбицкая А. Понимание культур через посредство ключевых слов. – М., 2001.
4. Зубов А.В., Зубова И.И. Информационные технологии в лингвистике. – М., 2004.
5. Кибрик А.Е. Очерки по общим и прикладным вопросам языкознания. – М., 2001.
6. Маслова В.А. Современные направления в лингвистике. – М., 2008.
7. Марчук Ю.Н. Основы компьютерной лингвистики. – М., 2000.
8. Машинный фонд русского языка: идеи и суждения. – М., 1986.
9. Нелюбин Л.Л. Компьютерная лингвистика и машинный перевод. – М., 2001.
10. Потапова Р.К. Новые информационные технологии и лингвистика. – М., 2012.
11. Hans Lindquist, *Corpus Linguistics and the Description of English*. Edinburg University Press, 2012.

7.3. Перечень ресурсов информационно-телекоммуникационной сети «Интернет»

1. Электронные библиотеки

1. [Российская Государственная Библиотека](http://www.rsl.ru/): <http://www.rsl.ru/>
2. Научная электронная библиотека: <http://txt.elibrary.ru/>
3. Научная библиотека Московского государственного университета им. М.В. Ломоносова: <http://www.lib.msu.ru/index.html>
4. Открытая русская электронная библиотека: <http://orel.rsl.ru/index.shtml>
5. Научная библиотека Санкт-Петербургского государственного университета: <http://www.lib.spbu.ru/>
6. Университетская информационная система «Россия»: <http://uisrussia.msu.ru>

2. Информационные ресурсы

1. Национальный корпус русского языка:
<http://www.ruscorpora.ru/>
2. Британский национальный корпус:
<http://www.natcorp.ox.ac.uk/>
3. Корпус современного американского английского:
<http://corpus.byu.edu/coca/>
4. AntConc:
<http://www.laurenceanthony.net/software.html>
руководство пользователя:
youtube.com/user/AntlabJPN
5. www.merriam-webster.com.
6. www.ozhegov.org.

8. Материально-техническое обеспечение

Учебная аудитория для проведения занятий лекционного и лабораторного типа, групповых и индивидуальных консультаций, текущего контроля и промежуточной аттестации, оснащенная следующим оборудованием: стандартная учебная мебель (28 учебных посадочных мест), стол и стул для преподавателя – по 1 шт., кафедра для лектора – 1 шт., доска настенная трехэлементная – 1 шт., напольный мобильный проекционный экран DA-LITE – 1 шт., мультимедиапроектор BenQ – 1 шт., ноутбук Lenovo – 1шт., колонки Genius – 1 шт.

9. Программное обеспечение

Программное обеспечение: Microsoft Open License (Windows XP, 7, 8, 10, Server, Office 2003-2016), лицензия 66975477 от 03.06.2016 (бессрочно).

**ДОКУМЕНТ ПОДПИСАН
ЭЛЕКТРОННОЙ ПОДПИСЬЮ**

Сертификат: 03B6A3C600B7ADA9B742A1E041DE7D81B0
Владелец: Артеменков Михаил Николаевич
Действителен: с 04.10.2021 до 07.10.2022