****

**1. Цели и результаты обучения дисциплины**

Рабочая программа дисциплины – программа освоения учебного материала, соответствующая требованиям государственного образовательного стандарта высшего профессионального образования и учитывающая специфику подготовки студентов по направлению или специальности.

Целями освоения дисциплины “Компьютерная лингвистика” являются знакомство с основными проблемами в области компьютерной лингвистики, базовыми алгоритмами, математическими методами моделирования языковых феноменов, основными инструментами и технологиями в области автоматической обработки естественного языка, умение представлять в алгоритмическом виде процессы анализа и синтеза текста.

**Место компьютерные технологии в лингвистике в структуре ООП «Переводческое дело» (колледж)**

Настоящая дисциплина относится к профессиональному блоку дисциплин «Компьютерная лингвистика».

Для специализаций Фундаментальная и компьютерная лингвистика профиля компьютерная лингвистика настоящая дисциплина является базовой.

Изучение данной дисциплины базируется на следующих дисциплинах:

* Введение в лингвистику (первый и второй курс) программы подготовки колледжа
* Теория языка
* Программирование и компьютерные инструменты лингвистического исследования

программы подготовки колледжа

* Программирование и компьютерные инструменты лингвистического исследования

программы подготовки колледжа

* Иностранный язык
1. **Результаты обучения и компетенции, формируемые в процессе изучения дисциплины**

|  |  |  |
| --- | --- | --- |
| **Код РО и его** **формулировка**  | **Код компетенции ООП и их формулировка** |  **РО дисциплины (РОд)**  **и их формулировка** |
| **РО 2**- Способен применять на практике теории перевода и понимать переводческие проблемы и осуществлять письменный и устный перевод, используя информационно коммуникативные технологии, математическому способами. | **ОК 5-**  Использует информационно-коммуникационные технологии в профессиональной деятельности;**ПК 14***-* Умеет использовать словари в печатной и электронной форме, используя информационные технологии**ДК-3**Использует современные информационные компьютерные технологии | Использовать информационно-коммуникационные технологии в профессиональной деятельности;Умеет использовать словари в печатной и электронной форме, используя информационные технологии Использует современные информационные компьютерные технологии |

**3.Место дисциплины в структуре ООП**

 Дисциплина “Компьютерные технологии в лингвистике” 2 курс (9 база) 1 курс (11-база) колледж относится к вариативной части цикла профессиональных дисциплин, согласно государственному образовательному стандарту базового профессионального образования по специальности: «Переводческое дело: английский язык». Данная дисциплина способствует освоению обучающимися лингвистических компонентов электронных информационных систем и служит расширению и углублению знаний студентов в области новых информационных технологий, а также развитию умения будущих лингвистов и переводчиков проводить экспертизу лингвистических программных продуктов и использовать их в дальнейшей учебной и научно-исследовательской и производственной деятельности. Необходимым требованием к «входным» знаниям, умениям и опыту деятельности обучающегося при освоении данной дисциплины является знакомство с фундаментальными основами и практикой использования средств информационно-коммуникационных технологий, таких как компьютер, средства связи, системное программное обеспечение, системы программирования, пакеты прикладных программ.

**4. Карта компетенций дисциплины в разрезе лекционных тем**

|  |  |  |  |
| --- | --- | --- | --- |
| **№** | **Разделы и названия темы** | **Кол.** **часов** | **Компетенции** |
| **ОК 5** | **ПК14** | **ДК3** | **Σ общее количество компетенций** |
| 1 | Предмет компьютерной лингвистики | 2ч | **+** |  | **+** | 2 |
| 2 | Обработка текстов и информационные технологии. | 2ч | **+** | **+** | **+** | 3 |
| 3 | Психологические аспекты компьютеризации в обучении  | 2ч |  | **+** |  | 1 |
| 4 | Общие принципы решения лингвистических задач методом моделирования. Основные этапы решения задач | 2ч | **+** | **+** |  | 2 |
| 5 | Машинный перевод текстов. Необходимость создания. Уровни автоматизации процесса перевода. Участие человека в переводе с использованием МП. Принципы работы, этапы. | 2ч | **+** | **+** | **+** | 3 |
| 6 | Интернет в работе переводчика. Возможности, ресурсы Способы и основные принципы повышения эффективности поиска в Интернете | 2ч | **+** |  | **+** | 2 |
|  | Всего | 12ч |  |  |  |  |

**Карта компетенций дисциплины в разрезе семинарских тем**

|  |  |  |  |
| --- | --- | --- | --- |
| **№** | **Разделы и названия темы** | **Кол.** **часов** | **Компетенции** |
| **ОК 5** | **ПК14** | **ДК3** | **Σ общее количество компетенций** |
| 1 | Предмет компьютерной лингвистики | 2ч | **+** |  | **+** | 2 |
| 2 | Обработка текстов и информационные технологии. | 2ч |  | **+** |  | 1 |
| 3 | Психологические аспекты компьютеризации в обучении  | 2ч | **+** | **+** | **+** | 3 |
| 4 | Общие принципы решения лингвистических задач методом моделирования. Основные этапы решения задач | 2ч | **+** |  | **+** | 2 |
| 5 | Машинный перевод текстов. Необходимость создания. Уровни автоматизации процесса перевода. Участие человека в переводе с использованием МП. Принципы работы, этапы. | 2ч | **+** | **+** |  | 2 |
| 6 | Интернет в работе переводчика. Возможности, ресурсы Способы и основные принципы повышения эффективности поиска в Интернете | 2ч | **+** |  | **+** | 2 |
| 7 | Автоматическое чтение тестов. Системы оптического распознавания символов и их возможности | 2ч | **+** | **+** | **+** | 3 |
| 8 | Translation Memory. Объективная необходимость создания. Принцип работы. | 2ч |  | **+** | **+** | 2 |
| 9 | Интернет в работе переводчика. Возможности, ресурсы | 2ч | **+** | **+** | **+** | 3 |
|  | Всего | 18ч |  |  |  |  |

**5. Технологическая карта дисциплины «Компьютерные технологии в лингвистике»**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **ТК** | **Разделы № и название темы** | **Количество часов** | **Сем** | **СРС** | **баллы** | **Общее количество компетенций** | **РК****10** |
| **Модуль 1**  |
| ТК 1 | Тема 1 | 2 | 2 | 2 | 0,3 | 1 |
| Тема 2 | 2 | 2 | 2 | 0,3 | 1 |
| ТК 2 | Тема 3 | 2 | 2 | 2 | 0,3 | 1 |
| Тема 4 | 2 | 2 | 2 | 0,3 | 1 |
| Тема 5 | 2 | 2 | 2 | 0,3 | 1 |
| ТК 3 | Тема 6 | 2 | 1,5 | 2 | 1.5 | 1 |
| Тема 7 | 2 | 1,5 | 2 | 1 | 1 |
| Тема 8 | 2 | 2 | 2 | 1 | 1 |
| **Модуль1** | **Всего** | 16 | 15 | 16 | 5 | 1 |
| **Модуль 2** | **РК****10** |
| ТК 4 | Тема 9  | 2 | 2 | 2 | 0,3 | 1 |
| Тема 10 | 2 | 2 | 2 | 0,3 | 1 |
| ТК 5 | Тема 11 | 2 | 2 | 2 | 0,3 | 1 |
| Тема 12 | 2 | 2 | 2 | 0,3 | 1 |
| Тема 13 | 2 | 2 | 2 | 0,3 | 1 |
| ТК 6 | Тема 14 | 2 | 2,5 | 2 | 1.5 | 1 |
| Тема 15 | 2 | 2,5 | 2 | 2 | 1 |
| **Модуль2** | **Всего** | 14 | 15 | 14 | 5 | 1 |

**6. Карта накопления баллов «Компьютерные технологии в лингвистике»**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Всего часов | Практ. (семин.) | СРС | **Модуль 1** | **Модуль 2** | **Экз** |
| Ауд.часы | СРС | Ауд.часы | СРС |  |
| Практика | Практика |
| 30 | 30 |  | 16ч | 16ч | 14ч | 14ч |  |
| Баллы | 15б | 5б | 15б | 5б |  |
| Итого модулей | К1 =10+15+5=30б | К2 =10+15+5=30б | К1+К2= 30+30+40 = 100б |

**7. Программа дисциплины**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Date**  | **Lesson**  | **Hours**  | **In-class activities(90 hours)**  | **Homework**  |
| **Sep.** | Lesson 11. The subject of computational linguistics • Introduction. The history of the origin of this item • Text analysis and its levels• Examples | 2 hours  | Dialogues Make up sentences.  | Read and supplement materials |
| **Sep.**  | Lesson 2 2. Text processing and information technology • General concept of text processing• Abstracting and examples | 2 hours  | Presentation. Dialogues.  | read and supplement materials |
| **Sep.**  | Lesson 3. Psychological aspects of computerization of training• Main directions | 2 hours  | Dialogues  | read and supplement materials |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Oct.** | **Lesson 4** General principles of solving linguistic problems by modeling. The main stages of solving the problem• Linguistic tasks • Modeling method • The main stages of solving the problem | 2 hours  | Presentation.  Make up sentencesDialogues  | read and supplement materials |
| **Oct.**  | Machine translation of texts. The need to create. Levels of automation of the translation process. Human participation in translation using MP. Principles of work, stages. | 2 hours | Presentation.  Make up sentences Dialogues  | read and supplement materials |
| **Oct.** | The Internet in the work of the translator. Opportunities, resources Ways and basic principles of improving the efficiency of Internet search | 2 hours | Presentation.  Make up sentences Dialogues  | read and supplement materials |
| **Oct.** | Automatic reading of tests. Optical character recognition systems and their capabilities | 2 hours | Presentation.  Make up sentences Dialogues  | read and supplement materials |
| **Oct.** | Translation Memory. Objective necessity of creation. The principle of operation. | 2 hours | Presentation.  Make up sentences Dialogues  | read and supplement materials |
| **Oct.** | The Internet in the work of the translator. Opportunities, resources | 2 hours | Presentation.  Make up sentences Dialogues  | read and supplement materials |

**8. Цели и результаты обучения по темам дисциплины**

|  |  |
| --- | --- |
| Темы  | РОд |
| 1. Предмет компьютерной лингвистики Введение. История происхождения данного предмета Анализ текста и его уровни | **Знать:** устройство компьютера, виды периферийных устройств и программного обеспечения;**Умеет:**основные направления использования информационных технологий в профессиональной деятельности лингвиста;особенности и принципы обработки русскоязычных и иноязычных текстов в производственно-практических целях с помощью компьютера;**Владеет:**методы использования программного обеспечения для решения конкретных лингвистических и переводческих задач и обучения иностранным языкам. |
| 2. Обработка текстов и информационные технологии Общее понятие об обработке текстовРеферирование и примеры  | **Знать:** устройство компьютера, виды периферийных устройств и программного обеспечения;**Уметь** использовать Future simple , future continuous , future perfect .**Владеет** : рассказывает о будущих планах , точно указывает будущее время действие . |
| 3. Психологические аспекты компьютеризации в обучении  | **Знать:** устройство компьютера, виды периферийных устройств и программного обеспечения;**Умеет** использовать структуры предложение и правильно составление вопросов всех видов .**Владеет** : правильно общение собеседником , конструктивное выражение своей речи  |
| 4. Общие принципы решения лингвистических задач методом моделирования. Основные этапы решения задач | **Знать:** устройство компьютера, виды периферийных устройств и программного обеспечения; интернет **Умеет** различать правильное использование методом моделирования.**Владеет :** рассказывать о своих возможностях и умений, использовать модальные глаголы в повседневной жизни в взаимодействии с друзьями, преподавателями. |
| 5. Машинный перевод текстов. Необходимость создания. Уровни автоматизации процесса перевода. Участие человека в переводе с использованием МП. Принципыработы, этапы. | **Знать:** устройство компьютера, виды периферийных устройств и программного обеспечения;**Умеет**: различать особенности использование артикулов, целенаправленное использование существительных и местоимений. **Владеет:** выражает себя свободно используя артикли, существительные и местоимении  |
| 6. Интернет в работе переводчика. Возможности, ресурсы Способы и основные принципы повышения эффективности поиска в Интернете | **Знает:**определения основным понятиям автоматической обработкитекста, воспроизводит базовые алгоритмы, используемы в автоматической обработке текста**Умеет:** основные пакеты морфологической обработки текста, демонстрирует знание базовых алгоритмов, владеет, использует современные методы тестирования качества**Владеет** :современные подходы к решениюзадач в области компьютернойлингвистики, интерпретирует результаты базовых алгоритмов |
| 7. Автоматическое чтение тестов. Системы оптического распознавания символов и их возможности | **Знает**: наиболее известные доступные для свободногоиспользования компоненты автоматического анализа, создавать модули первичной обработки текста**Умеет:** основные пакеты морфологической обработки текста, демонстрирует знание базовых алгоритмов, владеет, использует современные методы тестирования качества**Владеет** :современные подходы к решениюзадач в области компьютернойлингвистики, интерпретирует результаты базовых алгоритмов |
| 8. TranslationMemory. Объективная необходимость создания.Принцип работы. | **Знает:** рефлексировать (оценивать иперерабатывать) освоенные научные методы\* и способы деятельностиУмеет:разрабатывать и совершенствовать электронные языковые ресурсы, лингвистические компоненты информационных системВладеет:лингвистические модули компьютерных систем |
| 9. Интернет в работе переводчика. Возможности, ресурсы | **Знает:** произвести углубленный анализ проблем, постановку и обоснование проектных задач, связанныхс лингвистической деятельностью**Умеет:**основные пакеты морфологической обработки текста, демонстрирует знание базовых алгоритмов, владеет, использует современные методы тестирования качества**Владеет:**рассказывать о своих возможностях и умений, использовать модальные глаголы в повседневной жизни в взаимодействии с друзьями, преподавателями. |

9 Темы самостоятельной работы студента

1. Введение в компьютерную лингвистику. Задачи компьютерной лингвистики. Лингвистические системы. Этапы автоматической обработки текста. Поиск, индексация, оценка качества
2. Первичная обработка текста. Методы идентификации языка. Методы и проблемы сегментации. Методы сегментации без пробельных языков.
3. Методы дизамбигуации.машинное обучение: rule-induction, HMM, МaxЕnt, CRF
4. История компьютерной лингвистики.
5. Компьютерные методы лингвистических исследований.
6. Автоматический анализ текста.
7. Лингвистические модели.
8. Формализация языковой структуры
9. Компьютерная лексикография, электронные словари
10. Квантитативная лингвистика. Частотные словари.
11. Корпусная лингвистика
12. Использование корпусов текстов в научных исследованиях
13. Системы машинного перевода
14. Локализация электронных ресурсов и программного обеспечения
15. Лингвистические ресурсы и поиск в Интернет
16. Лингвистические технологии информационного поиска
17. Лингвистические аспекты искусственного интеллекта
18. Компьютерные технологии в изучении иностранных языков
19. Особенности научно-исследовательской и прикладной деятельности в области лингвистики с помощью компьютерных технологий
20. Теоретические, методологические и практические основы использования компьютерных технологий в лингвистике.
21. Автоматический синтаксический анализ. Анализ в терминах зависимостей
22. Алгоритмы анализа в терминах деревьев зависимостей
23. Психологические аспекты компьютеризации в обучении
24. Распознаватель языка / спелл-чеккер / токенизатор
25. Формализмы и методы автоматического синтаксического анализа

**10. Критерии оценки знаний, навыков**

 Данный курс в качестве текущего контроля предполагает выполнение практических заданий на семинарах и практических занятиях, а также двух домашних практических задания, результаты которых представляются студентами в виде проектов на языке Python, а также технических отчетов. Кроме того, на практических занятиях осуществляется постоянный текущий контроль в форме заданий, тестов, а также обсуждения текущих домашних упражнений.

 При выполнении домашнего задания студент должен продемонстрировать знание основных проблем и принципов первичной обработки текста, распознавания языка, на котором написан текст, методов разбиение текста на слова в безпробельных языках, умение работать с основными корпусами текстов на русском языке, знание основных свободно-распространяемых систем

морфологического анализа текста, умение запускать одну из систем, знание основных принципов анализа результатов морфологического анализа и принципов оценки качества морфологического анализа, умение анализировать результаты морфологического анализа, знание методов автоматического синтаксического анализа.

 При выполнении тестового задания на экзамене студенты должны продемонстрировать знание классификации основных задач компьютерной лингвистики, знание основных моделей и алгоритмов автоматического морфологического и синтаксического анализа текста, решать простые упражнения на применение базовых алгоритмов, быть в состоянии протестировать и оценивать

 работу отдельных модулей.

 Оценки по всем формам текущего контроля выставляются по 10-ти балльной шкале. Домашнее задание по подготовке корпуса, разработке программы предобработки текста/распознавания языка/исправления опечаток, разработке морфологических анализаторов являются групповыми проектами с индивидуальными заданиями, выполняемыми в рамках проектов. программы предварительного анализа текста и морфологических анализаторов проходит в

формате Форума по оценке систем автоматической обработке текста. Командам выдается тестовый и эталонный корпус. Каждая команда проводит оценку точности и полноты, а такжефункциональное тестирование соответствующей программы.

**11.Образовательные технологии**

Принцип функционально-коммуникативной направленности. Обучение должно строиться таким образом, чтобы вовлекать учащихся в устную (аудирование, говорение) и письменную (чтение, письмо) коммуникацию, т.е.общение на изучаемом языке на протяжении всего курса. При обучении слушанию, говорению и чтению необходимо учитывать специфику функционирования каждого вида речевой деятельности и языкового явления в процессе общения. Функциональность проявляется в организации языкового материала, в характере устно-речевых упражнений, в организации учебной работы на уроке.

1. Принцип дифференцированого и интегрированного обучения. Необходимо проводить разграничение в обучении устной и письменной речи; аудированию и говорению; чтению и письму, подготовленной и неподготовленной речи, поскольку в обучении каждому из указанных видов речи, форм речи решаются свои методические задачи и используются упражнения, адекватные формируемой деятельности, выполнение которых должно вести к овладению требуемыми навыками и умениями.
2. Принцип учета родного языка. Необходимо учитывать речевой опыт учащихся в родном языке при отборе содержания обучения, для прогнозирования трудностей при обучении произносительной, грамматической и лексической сторонам изучаемого языка, при обучении графике и орфографии, при семантизации, для определения структуры речевых действий.
3. Принцип устного опережения. Этим принципом определяется последовательность включения языкового материала в учебный процесс: от слушания и говорения к чтению и письму.
4. Принцип структурного подхода к освоению грамматического строя языка. Все грамматическое многообразие предложений сводится к ограниченному числу грамматических структур (типовых предложений).
5. Принцип личностно - ориентированного подхода, при котором учитываются возрастные особенности учащихся на каждой ступени обучения, их интересы, возможности и потребности, где деятельностная составляющая доминирует над знаниевой.
6. Принцип учета межпредметных связей. Межпредметные связи играют важную роль в повышении практической и научно-теоретической подготовки учащихся, существенной особенностью которой является овладением учащимися 3-11 классов обобщенным характером познавательной деятельности, что позволяет им применять приобретенные знания и умения в конкретных ситуациях повседневной, научной и общественной жизни.В условиях использования межпредметных связей на уроках иностранного языка учащиеся понимают необходимость применения знаний из других предметов, производят отбор нужных знаний из других предметов, переносят их в новую ситуацию, сопоставляют знания из смежных предметов, синтезируют знания и получают результат. Например, обладая определенными знаниями по истории, биологии и географии Кыргызстана, полученными на уроках истории, биологии и географии, ученики могут рассказать о своей стране (городе) на иностранном языке. При изучении иностранного языка учащиеся переносят знания и умения, полученные при изучении родного языка. В частности, это касается изучения грамматики, лексики, синтаксиса и пунктуации.

Представленные ниже методы обеспечивают реализацию заявленных принципов обучения иностранным языкам и ориентированы на овладение учащимися ключевыми и предметными компетентностями.

1. Прикладная лингвистика. // Статья в энциклопедии «Фонд знаний «Ломоносов»». URL:http://www.lomonosov-fund.ru/enc/ru/encyclopedia:01206:article

2. [J&M] - Introduction // Daniel Jurafsky& James H. Martin. Speech and Language Processing: An introduction to natural language processing, computational linguistics, and speech recognition. Second edition. 2009.http://stp.lingfil.uu.se/~santinim/ml/2014/JurafskyMartinSpeechAndLanguageProcessing2ed\_draft%202007.pdf

3. Большакова и др. (2011). Автоматическая обработка текстов на естественном языке и компьютерная лингвистика : Часть 2, глава 5-6.

*Дополнительная литература* Генерация текстов Соколова Е.Г., Болдасов М. Автоматическая генерация текстов на ЕЯ (портрет направления) <http://www.dialog-21.ru/Archive/2004/Sokolova.htm>EhudReiter. Has a Consensus NL Generation Structure Appeared, and is it Psycholinguistically Plausible? — 7th International Generation Workshop (Kennebunkport, Maine). URL:http://www.aclweb.org/anthology/W/W94/W94-0319.pdf

*Машинный перевод* Лекция Л.Л.Иомдина “Машинный перевод: успехи, неудачи, надежды”. Лекторий Политехнического музея. Видео. Доступно с URL http://rutube.ru/video/828268c50a49b876a6f4676b839fa745/ дата обращения 20.01.2022) Brown P. F. etal. The mathematics of statistical machine translation: Parameter estimation //Computational linguistics. – 1993. – Т. 19. – №. 2. – С. 263-311. http://acl.ldc.upenn.edu/J/J93/J93-2003.pdf

*Textmining*(классификация, кластеризация, реферирование). На примере анализа новостного потока Кондратьев М. Е. Анализ методов кластеризации новостного потока //Тр. Восьмой Всерос. науч. конф.«Электронные библиотеки: перспективные методы и технологии, электронные коллекции»(RCDL’2006).—Ярославль. – 2006. – С. 108-114.rcdl.ru/doc/2006/paper\_92\_v1.pdf

**12. Политик курса**

* + - * Не опаздывать на занятие
* Не разговаривать во время занятий, не жевать резинку
* Отключить сотовый телефон или переставить на беззвучный режим
* Не пропускать занятия, в случае болезни предоставить справку
* Пропущенные занятия отрабатывать в отведенное время

**13. Политика выставление баллов**

В соответствии с картой накопления баллов, студент может набирать баллы по всем видам занятий. На практических занятиях за квизы, диктант, выполнение упражнений, сочинений, конспектов студенты могут получить максимум 10 баллов; за выполнение домашних заданий, эссе, презентации, топики, а также за внеаудиторную работу получают максимум 10 баллов; за рубежный контроль – максимум 40 балл

**Информация об оценке**

|  |  |  |  |
| --- | --- | --- | --- |
| Рейтинг(баллы) | Оценка по буквенной системе | Цифровой эквивалент оценки | Оценка по традиционной системе |
| 87 – 100 | A | 5 | Отлично |
| 74 -86 | B | 4 | Хорошо |
| 61 – 73 | C | 3 | Удовлетворительно |
| 31 – 60 | D | 2 | Неудовлетворительно |

Выставление оценок осуществляется на основе принципов объективности, справедливости, всестороннего анализа качества знаний студентов, и других положений, способствующих повышению надежности оценки знаний обучающихся и устранению субъективных факторов.