|  |  |
| --- | --- |
| Национальный  исследовательский университет  «Высшая школа экономики»  **Лицей** | **Приложение 433**  УТВЕРЖДЕНО  педагогическим советом  Лицея НИУ ВШЭ  протокол №14 от 24.08.2022г. |

**Рабочая программа учебного предмета (курса)**

«Фундаментальная и компьютерная лингвистика»

**10 класс**

**Автор:**

Виняр Алексей Игоревич

Попова Маргарита Владимировна

Казакова Татьяна Борисовна

**1. Планируемые результаты освоения учебного предмета**

В соответствии с Федеральным государственным образовательным стандартом среднего общего образования (10-11 кл.) освоение учебного предмета «Фундаментальная и компьютерная лингвистика» предполагает достижение личностных, метапредметных и предметных результатов.

Личностные результаты освоения учебного предмета включают в себя:

* формирование ответственного отношения к учению, готовности и способности обучающихся к саморазвитию и самообразованию на основе мотивации к обучению и познанию, осознанному выбору и построению дальнейшей индивидуальной траектории образования на базе ориентировки в мире профессий и профессиональных предпочтений, с учетом устойчивых познавательных интересов, а также на основе формирования уважительного отношения к труду, развития опыта участия в социально значимом труде;
* формирование целостного мировоззрения, соответствующего современному уровню развития науки и общественной практики, учитывающего социальное, культурное, языковое, духовное многообразие современного мира;
* формирование коммуникативной компетентности в общении и сотрудничестве со сверстниками, детьми старшего и младшего возраста, взрослыми в процессе образовательной, общественно полезной, учебно-исследовательской, творческой и других видов деятельности;

Предметные результаты освоения учебного предмета включают в себя:

* понимание главных положений теоретической грамматики  естественных языков;
* понимание основных задач математической лингвистики; овладение базовым формальным аппаратом описания языка
* формирование навыков проведения формального анализа грамматики языка;
* расширение и систематизацию научных знаний о языке; осознание взаимосвязи его уровней и единиц; освоение базовых понятий лингвистики, основных единиц и грамматических категорий языка;
* формирование навыков проведения различных видов анализа слова (фонетического, морфемного, словообразовательного, лексического, морфологического), синтаксического анализа словосочетания и предложения, а также многоаспектного анализа текста;
* обогащение активного и потенциального словарного запаса, расширение объема используемых в речи грамматических средств для свободного выражения мыслей и чувств адекватно ситуации и стилю общения

**2. Содержание учебного предмета**

**Тема 1. Лингвистические задачи**

Самодостаточные лингвистические задачи. Задачи, требующие владения одним или несколькими иностранными языками. Задачи с использованием материала одного или нескольких языков. Задачи на сравнение данных родственных языков. Задачи с перепутанными соответствиями.

**Тема 2. Лингвистические данные. Корпуса**

Типы лингвистических данных. Полевые данные. Словари. Грамматики. Корпуса (текстовые или мультимедийные).

Корпуса. Чем корпус отличается от набора текстов. Типы корпусов. Известные корпуса.

**Тема 3. Национальный корпус русского языка**

Как устроен поиск в НКРЯ.

Обзор подкорпусов НКРЯ

Типы исследований на базе НКРЯ (практическое занятие).

Анализ проблем. Морфологическая и лексическая омонимия Дисбалансы в корпусе.

**Тема 4. Частотность**

Что может сказать о тексте количество употреблений определенного слова.

Частотности единиц текста и их важность для лингвистических исследований

Tf/idf. Извлечение ключевых слова из текста.

**Тема 5. Интернет как лингвистический корпус для исследования**

Что можно исследовать прямо в поисковике. Синтаксис и возможности расширенного поискового запроса.

Веб как спеллчекер.

Говорят ли так? Веб как помощник переводчика (помимо Google Translate)

Исследование региональных вариаций языка с помощью расширенного запроса

Веб как источник знаний о мире (методы майнинга структурированной информации)

ГИКРЯ и его возможности.

**Тема 6. Конечные автоматы и регулярные выражения**

**Тема 7. Морфологический анализ**

Морфология как раздел науки о языке. Предмет морфологии.

Грамматическое vs лексическое значение. Словоизменительная парадигма. Словарь Зализняка.

Стеммеры. Стеммер Портера.

Морфологические анализаторы. Mystem.

Морфологическая омонимия.

Морфологическая разметка в НКРЯ и других корпусах.

**Тема 8. Синтаксический анализ**

Синтаксис как раздел науки о языке. Предмет синтаксиса.

Варианты иерархической связи между единицами речи. Линейный порядок. Древесная структура.

Синтаксические формализмы. Структура зависимостей. Структура составляющих.

Синтаксические деревья. Синтаксическая разметка в НКРЯ и других корпусах.

**Тема 9. Лексикографические ресурсы и их применение в компьютерной лингвистике/автоматическом анализе текста**

Традиционные лексикографические ресурсы (словари).

Тезаурусы и лексические онтологии. Рутез, ворднет.

Меры семантической близости слов.

**Тема 10. Введение в статистические подходы к анализу естественного языка.**

Поверхностное объяснение сути машинного обучения («обучение на примерах»).

Демонстрация применимости машинного обучения к задачам комп.лингвистики.

Отбор релевантных признаков для задачи классификации текстов.

**Тема 11. Задача извлечения информации из текста**

История. DARPA и MUC.

Простые инструменты извлечения информации с опорой на морфологию. Tomita парсер. GATE.

**Тема 12. Машинный перевод**

История развития машинного перевода. Холодная война, Джорджтаунский эксперимент, ALPAC.

Правиловый машинный перевод.

Статистический машинный перевод, параллельные корпуса, выравнивание.

Тестирование доступных систем MП и анализ ошибок.

**Тема 13. Анализ тональности текста**

Как мы понимаем, что в тексте хвалят или ругают что-то или кого-то. Маркеры тональности. Инверторы и интенсификаторы.

Как это может понять компьютер.

Тестирование систем, анализ ошибок, выводы о подводных камнях.

**Тема 14. Разрешение анафоры**

Референция. Анафорическая связь. Антецедент. Анафор. Катафора.

**Тема 15. Распознавание речи.**

Физические свойства звукового сигнала и их математическая формализация

Лингвистика в распознавании речи.

**Тема 16. Неоднозначность в языке.**

Разные уровни неоднозначности. Морфологическая, лексическая, синтаксическая

Способы моделирования значений и снятия неоднозначности.

**Тема 17. Компьютерная лингвистика и педагогика**

Применение методов автоматической обработки языка в образовании.

**Тема 18. Компьютерная лингвистика и литература**

Автопоэты и автописатели. Цифровое литературоведение и "дальнее чтение". Компьютерная стилистика. Исследование: сравниваем язык Пушкина с языком Гоголя. Определение авторства текста.

**3. Тематическое планирование**

|  |  |  |  |
| --- | --- | --- | --- |
| № | Тема урока | Количество аудиторных часов | Основные виды деятельности |
| 1 | Лингвистические задачи | 4 | Лекция и дискуссия по обозначаемой преподавателем проблематике. Решение задач. |
| 2 | Лингвистические данные. Корпуса | 4 | Лекция и дискуссия по обозначаемой преподавателем проблематике. Практикум |
| 3 | Национальный корпус русского языка | 2 | Лекция и дискуссия по обозначаемой преподавателем проблематике. Практикум. |
| 4 | Частотность. Частотности единиц текста и их важность для лингвистических исследований | 2 | Лекция и дискуссия по обозначаемой преподавателем проблематике. Практикум. В начале урока - небольшая контрольная по теме "лингвистические корпуса" |
| 5 | Интернет как лингвистический корпус для исследования. | 4 | Лекция и дискуссия по обозначаемой преподавателем проблематике. Практикум. |
| 6 | Морфологический анализ | 6 | Лекция и дискуссия по обозначаемой преподавателем проблематике. Практикум. |
| 7 | Конечные автоматы и регулярные выражения. | 4 | Лекция и дискуссия по обозначаемой преподавателем проблематике. Практикум. |
| 8 | Синтаксический анализ | 6 | Лекция и дискуссия по обозначаемой преподавателем проблематике. Практикум. Задание по построению синтаксических деревьев |
| 9 | Лексикографичес-кие ресурсы и их применение в компьютерной лингвистике/автоматическом анализе текста | 2 | Лекция и дискуссия по обозначаемой преподавателем проблематике. Практикум. |
| 10 | Введение в статистические подходы к анализу естественного языка. | 6 | Лекция и дискуссия по обозначаемой преподавателем проблематике. Практикум. |
| 11 | Задача извлечения информации из текста | 3 | Лекция и дискуссия по обозначаемой преподавателем проблематике. Практикум. В начале первого урока - небольшая контрольная-опрос по базовым понятиям морфологии/синтаксиса |
| 12 | Машинный перевод | 3 | Лекция и дискуссия по обозначаемой преподавателем проблематике. Практикум. |
| 13 | Анализ тональности текста | 3 | Лекция и дискуссия по обозначаемой преподавателем проблематике. Практикум. |
| 14 | Разрешение анафоры | 3 | Лекция и дискуссия по обозначаемой преподавателем проблематике. Практикум. |
| 15 | Распознавание речи | 3 | Лекция и дискуссия по обозначаемой преподавателем проблематике. Практикум. |
| 16 | Неоднозначность в языке | 3 | Лекция и дискуссия по обозначаемой преподавателем проблематике. Практикум. Письменное задание в классе на приведение примеров неоднозначности разных типов. |
| 17 | Компьютерная лингвистика и педагогика | 2 | Лекция и дискуссия по обозначаемой преподавателем проблематике. Практикум. |
| 18 | Компьютерная лингвистика и литература. | 6 | Лекция и дискуссия по обозначаемой преподавателем проблематике. Практикум. |
|  |  | 66 |  |

В воспитании обучающихся приоритетом является создание благоприятных условий для приобретения обучающимися опыта осуществления социально значимых дел:

— опыт дел, направленных на заботу о своей семье, родных и близких;

— трудовой опыт, опыт участия в производственной практике;

— опыт дел, направленных на пользу своему родному городу или селу, стране   
в целом, опыт деятельного выражения собственной гражданской позиции;

— опыт природоохранных дел;

— опыт разрешения возникающих конфликтных ситуаций в школе, дома   
или на улице;

— опыт самостоятельного приобретения новых знаний, проведения научных исследований, опыт проектной деятельности;

— опыт изучения, защиты и восстановления культурного наследия человечества, опыт создания собственных произведений культуры, опыт творческого самовыражения;

— опыт ведения здорового образа жизни и заботы о здоровье других людей;

— опыт оказания помощи окружающим, заботы о малышах или пожилых людях, волонтерский опыт;

— опыт самопознания и самоанализа, опыт социально приемлемого самовыражения и самореализации.

Выделение данного приоритета связано с особенностями обучающихся юношеского возраста: с их потребностью в жизненном самоопределении, в выборе дальнейшего жизненного пути, который открывается перед ними на пороге самостоятельной взрослой жизни.

**Учебно-методическое обеспечение образовательной деятельности:**

1. Львова С.И., Львов В.В. Русский язык (базовый и углублённый уровни). 10 класс. М:. ООО «ИОЦ МНЕМОЗИНА»
2. Хомский Н., Синтаксические структуры, в сб.: «Новое в лингвистике», в. 2, М., 1962.
3. Гладкий А. В., Мельчук И. А., Элементы математической лингвистики, М., 1969.
4. Зализняк А.А. Русское именное словоизменение, М., 1967.
5. Плунгян, В. А. Общая морфология: Введение в проблематику. 2012.