

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ
РОССИЙСКОЙ ФЕДЕРАЦИИ
Федеральное государственное автономное образовательное учреждение
высшего образования
«ТЮМЕНСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ»
Тобольский педагогический институт им. Д.И.Менделеева
(филиал) Тюменского государственного университета

21

УТВЕРЖДАЮ
Директор
Шидлов С.П.
« 28 » _____ 2020 г.



ВВЕДЕНИЕ В НЕЙРОННЫЕ СЕТИ И МАШИННЫЙ ПЕРЕВОД
Рабочая программа
для обучающихся по направлению подготовки
45.03.02 ЛИНГВИСТИКА
Профиль Теория и методика преподавания иностранных языков и культур
Форма обучения: очная

Шитиков Петр Михайлович. Введение в нейронные сети и машинный перевод. Рабочая программа для обучающихся по направлению подготовки 45.03.02 Лингвистика, профиль Теория и методика преподавания иностранных языков и культур, форма обучения очная. Тобольск, 2020.

Рабочая программа дисциплины (модуля) опубликована на сайте ТюмГУ: Введение в нейронные сети и машинный перевод [электронный ресурс] / Режим доступа: <https://tobolsk.utmn.ru/sveden/education/#>

21

© Тобольский педагогический институт им. Д.И. Менделеева (филиал) Тюменского Государственного университета, 2020

© Шитиков Петр Михайлович, 2020

1. Пояснительная записка

Цель изучения курса «Введение в нейронные сети и машинный перевод» состоит в формировании у студентов знаний о принципах и методах разработки систем машинного перевода, об их лингвистическом обеспечении, об истории разработок и существующих системах, а также в формировании умений и навыков формализации знаний о языке и алгоритмизации лингвистического анализа с целью создания лингвистического процессора систем машинного перевода.

Задачи курса:

- ознакомить студентов с теоретическими основами, методами и средствами формализации языка;
- выработать умения составления лингвистических алгоритмов анализа слова и его значения, структуры и семантики предложения и текста при создании лингвистического процессора системы машинного перевода;
- сформировать навыки использования систем машинного перевода

1.1. Место дисциплины (модуля) в структуре образовательной программы

Данная дисциплина относится к циклу Б1. Дисциплины (модули), к дисциплинам по выбору студента вариативной части ОП.

Курс предполагает углубленное знакомство с современными направлениями использования языка в информационных технологиях; с формализацией знаний о языке и алгоритмизацией лингвистического анализа; с принципами создания лингвистических банков данных и баз знаний. Приложение знаний специализации может найти применение в преподавании иностранного языка с использованием новых информационных технологий в школе и вузе; в создании систем машинного перевода; в создании лингвистических банков данных и научной работе на базе этих банков данных.

Данная учебная дисциплина логически взаимосвязана с такими дисциплинами как «Основы корпусной лингвистики», «Информационные технологии в преподавании иностранного языка», «Методы лингвистического исследования», «Актуальные проблемы прикладной лингвистики». Она призвана сформировать у обучающихся систему знаний, умений и готовностей в сфере современных лингвистических информационных технологий в контексте их использования в филологии и образовании.

Знания, полученные студентами в ходе изучения курса, будут необходимы при подготовке к преддипломной практике, а также к государственной итоговой аттестации и в процессе ее.

1.2. Компетенции обучающегося, формируемые в результате освоения данной дисциплины (модуля).

Код и наименование компетенции (из ФГОС ВО)	Компонент (знаниевый/функциональный)
ПК-23 способность использовать понятийный аппарат философии, теоретической и прикладной лингвистики, переводоведения, лингводидактики и теории межкультурной коммуникации для	Знает сведения о круге проблем, составляющих в настоящее время предмет изучения в области теории языка и языковой деятельности, а также в сопряженных науках (нейролингвистике и переводоведении)
	Умеет устанавливать межпредметные связи и

решения профессиональных задач	использовать знания, полученные в других гуманитарных науках, для осмысления, объяснения и интерпретации лингвистических (переводоведческих) проблем
ПК-25 владение основами современных методов научного исследования, информационной и библиографической культурой	Знает методологию лингвистического исследования в области нейронных сетей и машинного перевода
	Умеет использовать методологию лингвистического исследования в области нейронных сетей и машинного перевода
ПК-26 владение стандартными методиками поиска, анализа и обработки материала исследования	Знает стандартные методики поиска, анализа и обработки материалов лингвистического (переводоведческого) исследования
	Умеет осмысленно выбирать научный метод для своего исследования в области лингвистического (переводоведческого) исследования

2. Структура и объем дисциплины

Таблица 1

Вид учебной работы	Всего часов (академические часы)	Часов в семестре (академические часы)
		7
Общий объем зач. ед.	4	4
	144	144
Часы аудиторной работы (всего):		
Лекции	16	16
Практические занятия		
Лабораторные / практические занятия по подгруппам	48	48
Часы внеаудиторной работы, включая самостоятельную работу обучающегося	80	80
Вид промежуточной аттестации (зачет, диф. зачет, экзамен)		зачет

3. Система оценивания

Оценивание осуществляется в рамках балльной системы, разработанной преподавателем и доведенной до сведения обучающихся на первом занятии

№	Виды оцениваемой работы	Количество баллов	
		Текущий контроль	Промежуточный контроль
1.	Посещение занятия	0/2	-

2.	Реферат	0-8	-
3.	Устный ответ	0-5	0-30

Промежуточная аттестация может быть выставлена по совокупности баллов, полученных обучающимся в рамках текущего контроля.

Перевод баллов в оценки (зачет)

№	Баллы	Оценки
1.	0-60	Не зачтено
2.	61-100	Зачтено

4. Содержание дисциплины

4.1. Тематический план дисциплины

Таблица 2

№ п/п	Наименование тем и/или разделов	Объем дисциплины (модуля), час.				
		Всего	Виды аудиторной работы (академические часы)			Иные виды контактно й работы
			Лекции	Практическ ие занятия	Лабораторны е/ практические занятия по подгруппам	
1	2	3	4	5	6	7
1.	Компьютерные технологии в филологии. Введение в проблематику машинного перевода.	18	2		2	-
2.	Базовая архитектура нейронных сетей	18	2		6	-
3.	Машинный перевод как направление искусственного интеллекта	18	2		6	-
4.	Виды и стратегии машинного перевода	18	2		6	-
5.	Алгоритм машинного перевода, основанного на лингвистическом анализе	18	2		6	-
6.	Структура систем машинного перевода	18	2		8	-
7.	Лингвистические проблемы машинного перевода	18	2		8	-
8.	Использование параллельных корпусов текстов в машинном переводе. Память переводов	18	2		8	-
	Итого (часов)	144	16		48	-

4.2. Содержание дисциплины (модуля) по темам

4.2.1 Лекционный курс

Компьютерные технологии в филологии. Введение в проблематику машинного перевода.	<ol style="list-style-type: none">1. Предмет и задачи машинного перевода2. Филологические направления, в которых активно задействуются компьютерные технологии.3. Экскурс в проблемы компьютерного перевода естественно-языковых текстов
Базовая архитектура нейронных сетей	<ol style="list-style-type: none">1. Одиночный вычислительный слой: перцептрон2. Взаимосвязь с методом опорных векторов3. Выбор функции активации и функции потерь4. Многослойные нейронные сети
Машинный перевод как направление искусственного интеллекта	<ol style="list-style-type: none">1. Определения машинного перевода.2. Краткие сведения из истории развития направления.3. Стимулы к развитию исследований.4. Преимущества машинного перевода.
Виды и стратегии машинного перевода	<ol style="list-style-type: none">1. Классификация переводов.2. Типы машинного перевода по степени автоматизации.3. Стратегии машинного перевода.
Алгоритм машинного перевода, основанного на лингвистическом анализе	<p>Шаг 1. Получение предложения исходного текста из файла или из буфера в памяти.</p> <p>Шаг 2. Разбиение предложения на слова и определение границ предложения.</p> <p>Шаг 3. Морфологический анализ исходного текста – получение всех возможных лексических кодов для каждого найденного в словаре слова.</p> <p>Шаг 4. Синтаксический анализ исходного текста – группировка однородных прилагательных и существительных, построение дерева главных/зависимых слов.</p> <p>Шаг 5. Семантический анализ исходного текста.</p> <p>Шаг 6. Осуществление перевода построенного дерева.</p> <p>Шаг 7. Осуществление согласования переведенного дерева – семантический, синтаксический и морфологический синтез.</p> <p>Шаг 8. Запись переведённого предложения в файл или в буфер</p>
Структура систем машинного перевода	<ol style="list-style-type: none">1. Вспомогательные программные средства.2. Состав лингвистической базы данных.3. Лингвистический процессор.
Лингвистические проблемы машинного перевода	<ol style="list-style-type: none">1. Проблема многозначности при машинном переводе.2. Синтаксические трансформации в машинном переводе.3. Перевод фразеологических сочетаний в системах машинного перевода.
Использование	<ol style="list-style-type: none">1. Виды и структура параллельных корпусов текстов.

параллельных корпусов текстов в машинном переводе. Память переводов	2. Технология памяти переводов. 3. Преимущества и недостатки технологии.
---	---

Лабораторные занятия

1. Компьютерные технологии в филологии. Введение в проблематику машинного перевода.
2. Базовая архитектура нейронных сетей
3. Машинный перевод как направление искусственного интеллекта
4. Виды и стратегии машинного перевода
5. Алгоритм машинного перевода, основанного на лингвистическом анализе
6. Структура систем машинного перевода
7. Лингвистические проблемы машинного перевода
8. Использование параллельных корпусов текстов в машинном переводе. Память переводов

5. Учебно-методическое обеспечение и планирование самостоятельной работы обучающихся

Таблица 3

№ темы	Темы	Формы СРС, включая требования к подготовке к занятиям
1.	Компьютерные технологии в филологии. Введение в проблематику машинного перевода.	Конспектирование первоисточников, изучение научной, учебно-методической литературы
2.	Базовая архитектура нейронных сетей	Доработка конспектов лекций: подбор, изучение, анализ и конспектирование рекомендованной литературы.
3.	Машинный перевод как направление искусственного интеллекта	Самостоятельное изучение отдельных тем учебной программы (или отдельных вопросов, проблем темы) с последующим контролем выполнения задания: предоставлением рефератов, комментариев или в виде устного ответа
4.	Виды и стратегии машинного перевода	Работа с литературой по заданной теме, составление вопросов для анализа, опорных конспектов
5.	Алгоритм машинного перевода, основанного на лингвистическом анализе	Самостоятельное изучение отдельных тем учебной программы (или отдельных вопросов, проблем темы) с последующим контролем выполнения задания: предоставлением рефератов, комментариев или в виде устного ответа
6.	Структура систем машинного перевода	Конспектирование первоисточников, изучение научной, учебно-методической литературы
7.	Лингвистические проблемы машинного перевода	Предоставление реферата по заданной теме

8.	Использование параллельных корпусов текстов в машинном переводе. Память переводов	Подготовка презентаций
----	---	------------------------

6. Промежуточная аттестация по дисциплине (модулю)

6.1 Оценочные материалы для проведения промежуточной аттестации по дисциплине

Примерные вопросы к зачету

1. Машинный перевод как научное направление, цели и задачи. Предмет и объект данной области знаний.
2. История машинного перевода. Три поколения развития.
3. Преимущества машинного перевода и цели разработки.
4. Виды, формы, типы ручного перевода. Типология перевода в зависимости от степени автоматизации: ручной, автоматизированный, автоматический перевод.
5. Стратегии машинного перевода: пословный перевод, перевод с использованием трансфера, с использованием интерлингвы, перевод, основанный на глубинном лингвистическом анализе, память переводов.
6. Общий алгоритм решения задачи машинного перевода.
7. Синтез текста на переводном языке: семантическое представление текста, построение синтаксических структур, морфологический синтез.
8. Структура лингвистической базы знаний систем машинного перевода: лингвистическая база данных и лингвистический процессор.
9. Компоненты лингвистической базы данных систем машинного перевода.
10. Морфологический словарь в структуре лингвистической базы данных.
11. Синтаксический словарь в структуре лингвистической базы данных.
12. Семантический словарь в структуре лингвистической базы данных.
13. Структура лингвистического процессора систем машинного перевода.
14. Морфологический и синтаксический анализ исходного текста в системах машинного перевода.
15. Семантический и прагматический анализ исходного текста в системах машинного перевода.
16. Лингвистические проблемы машинного перевода на морфологическом уровне. Пути решения проблем.
17. Лингвистические проблемы машинного перевода на синтаксическом уровне. Пути решения проблем.
18. Лингвистические проблемы машинного перевода на семантическом уровне. Пути решения проблем.
19. Лингвистические проблемы машинного перевода на прагматическом уровне. Пути решения проблем.
20. Синтаксические трансформации в машинном переводе.
21. Перевод фразеологических сочетаний в машинном переводе.
22. Учет контекста и экстралингвистических факторов при разрешении полисемии.
23. Системы машинного перевода, основанные на примерах. Принципы построения корпуса текстов для систем памяти переводов.

24. Структура систем и алгоритм работы памяти переводов. Сфера применения.
25. Контекстологический словарь в структуре систем машинного перевода.
26. Критерии оценки качества машинного перевода. Оценка перевода испытуемыми. Шкала оценки качества перевода.
27. Критерии оценки качества машинного перевода. Статистический критерий. Изучение ответов по оригиналу текста человека, читавшего только перевод.
28. Оценка качества по трудозатратам на корректировку машинного перевода. Качество перевода на различных языковых уровнях.
29. Машинный перевод в контексте исследований в области искусственного интеллекта. Машинный перевод и компьютерная лингвистика.
30. Машинный перевод и информационный поиск.
31. Связь машинного перевода с математической лингвистикой и психолингвистикой.
32. Машинный перевод и теория перевода.
33. Машинный перевод в деятельности переводчика.

Примерные темы рефератов

1. Машинный перевод в контексте исследований в области искусственного интеллекта.
2. Машинный перевод и компьютерная лингвистика.
3. Связь машинного перевода с математической лингвистикой и психолингвистикой.

6.2 Критерии оценивания компетенций:

Таблица 4

Карта критериев оценивания компетенций

№ п/п	Код и наименование компетенции	Компонент (знаниевый/функциональный)	Оценочные материалы	Критерии оценивания
1	ПК-23 способность использовать понятийный аппарат философии, теоретической и прикладной лингвистики, переводоведения, лингводидактики и теории межкультурной коммуникации для решения профессиональных задач	Знает сведения о круге проблем, составляющих настоящее время предмет изучения в области теории языка и языковой деятельности, а также в сопряженных науках (нейролингвистике и переводоведении) Умеет устанавливать межпредметные связи и	Реферат Экзаменационные вопросы	1. Знает предмет изучения в области теории языка и языковой деятельности, а также в сопряженных науках (нейролингвистике и переводоведении) 2. Умеет знания, полученные в других гуманитарных науках, для осмысления, объяснения и интерпретации лингвистических

		использовать знания, полученные в других гуманитарных науках, для осмысления, объяснения и интерпретации лингвистических (переводоведческих) проблем		(переводоведческих) проблем 3. Демонстрирует умение устанавливать межпредметные связи в области теории языка и языковой деятельности
2	ПК-25 владение основами современными методами научного исследования, информационной и библиографической культурой	Знает методологию лингвистического исследования в области нейронных сетей и машинного перевода Умеет использовать методологию лингвистического исследования в области нейронных сетей и машинного перевода	Реферат Экзаменационные вопросы	1. Знает основные методологические принципы исследования в области нейронных сетей и машинного перевода 2. Умеет применять изученные принципы исследования на практике 3. Демонстрирует навык осуществления лингвистического исследования на основании методологии
3	ПК-26 владение стандартными методиками поиска, анализа и обработки материала исследования	Знает стандартные методики поиска, анализа и обработки материалов лингвистического (переводоведческого) исследования Умеет осмысленно выбирать научный метод для своего исследования в области лингвистического (переводоведческого) исследования	Реферат Экзаменационные вопросы	1. Знает принципы поиска материалов лингвистического (переводоведческого) исследования 2. Умеет применять на практике принципы анализа и обработки материалов лингвистического 3. Демонстрирует навык осмысленного выбора научный метод для своего исследования

7. Учебно-методическое и информационное обеспечение дисциплины (модуля)

7.1 Основная литература:

1. **Пиванова, Э. В.** Теория и практика машинного перевода : учебное пособие / Э. В. Пиванова. — Ставрополь : Северо-Кавказский федеральный университет, 2014. — 115 с. —URL: <http://www.iprbookshop.ru/63021.html>

7.2 Дополнительная литература:

1. **Горожанина, Е. И.** Нейронные сети : учебное пособие / Е. И. Горожанина. — Самара : Поволжский государственный университет телекоммуникаций и информатики, 2017. — 84 с. — URL: <http://www.iprbookshop.ru/75391.html>

2. **Захарова, Т. В.** Практические основы компьютерных технологий в переводе : учебное пособие / Т. В. Захарова, Е. В. Турлова. — Оренбург : Оренбургский государственный университет, ЭБС АСВ, 2017. — 109 с. — URL: <http://www.iprbookshop.ru/71314.html>

7.3. Интернет-ресурсы: не предусмотрены

7.4. Современные профессиональные базы данных и информационные справочные системы

Знаниум - <https://new.znanium.com/>

Лань - <https://e.lanbook.com/>

IPR Books - <http://www.iprbookshop.ru/>

Elibrary - <https://www.elibrary.ru/>

Национальная электронная библиотека (НЭБ) - <https://rusneb.ru/>

Межвузовская электронная библиотека (МЭБ) - <https://icdlib.nspu.ru/>

"ИВИС" (БД периодических изданий) - <https://dlib.eastview.com/browse>

Электронная библиотека Тюмгу - <https://library.utmn.ru/>

8. Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине (модулю):

Платформа для электронного обучения MicrosoftTeams.

Бесплатное и условно-бесплатное программное обеспечение, установленное в аудиториях: 7-Zip, AdobeAcrobatReader, GIMP, ModelVisionStudium, GoogleChrome, MozillaFirefox, OpenOffice.org, UVScreenCamera, UVSoundRecorder, SMathStudioDesktop, Inkscape, MyTestX, WinVDIG, OracleVirtualBox, AdobeMediaPlayer, Kompozer.

Лицензионное программное обеспечение, установленное в аудиториях: Autodesk 3ds Max 2018, Autodesk AutoCAD 2018, Embarcadero RAD Studio 2010, MatLab R2009a, Microsoft Office 2003, Microsoft Office 2007, Microsoft Office 2010, Microsoft Visual Studio 2012, Microsoft Visual Studio 2012 Expression, Microsoft Visual FoxPro 9.0, Microsoft SQL Server 2005, Windows, Dr. Web, Конструктор тестов 2.5 (Keepsoft), Adobe Design Premium CS4.

9. Технические средства и материально-техническое обеспечение дисциплины (модуля)

- Учебные аудитории для проведения лекций и практических занятий укомплектованы специализированной мебелью и техническими средствами обучения, служащими для представления учебной информации большой аудитории: мультимедийные аудитории, укомплектованные таким оборудованием, как проектор, документ камера, проекционным экраном; лингафонный кабинет, насчитывающий 15 рабочих мест ученика и 1 рабочее место преподавателя.
- Помещения для самостоятельной работы обучающихся (компьютерные классы) оснащены компьютерной техникой с возможностью подключения к сети «Интернет» и обеспечены доступом к электронной информационно-образовательной среде вуза.
- Лаборатории, оснащенные лабораторным оборудованием, для реализации данной дисциплины не предусмотрены.