

Минобрнауки России

Федеральное государственное бюджетное образовательное учреждение
высшего образования

«Оренбургский государственный университет»

Кафедра теории и практики перевода

РАБОЧАЯ ПРОГРАММА

ДИСЦИПЛИНЫ

«Б.1.В.ДВ.3.2 Компьютерная обработка текста»

Уровень высшего образования

БАКАЛАВРИАТ

Направление подготовки

45.03.02 Лингвистика

(код и наименование направления подготовки)

Перевод и переводоведение (английский язык, второй иностранный язык)

(наименование направленности (профиля) образовательной программы)

Тип образовательной программы

Программа академического бакалавриата

Квалификация

Бакалавр

Форма обучения

Очная

Год набора 2016

Рабочая программа рассмотрена и утверждена на заседании кафедры

Кафедра теории и практики перевода

наименование кафедры

протокол № 7 от "19" 01 2016 г.

Заведующий кафедрой

Кафедра теории и практики перевода

наименование кафедры

подпись

А.В. Люлина

расшифровка подписи

Исполнители:

Старший преподаватель

должность

подпись

У.С. Баймуратова

расшифровка подписи

Доцент

должность

подпись

Е.Д. Андреева

расшифровка подписи

СОГЛАСОВАНО:

Председатель методической комиссии по направлению подготовки

45.03.02 Лингвистика

код наименование

подпись

И.А. Солодилова

расшифровка подписи

Заведующий отделом комплектования научной библиотеки

личная подпись

Н.Н. Грицай

расшифровка подписи

Уполномоченный по качеству факультета

личная подпись

Т.В. Сапух

расшифровка подписи

№ регистрации _____

1 Цели и задачи освоения дисциплины

Цель (цели) освоения дисциплины:

развитие способности применять автоматический лингвистический анализ и синтез текста, а также ознакомление студентов с лингвистическими аспектами общения с компьютером на естественном языке.

Задачи:

- овладеть основами автоматического индексирования и реферирования текстов, предредактирования и постредактирования, статистической обработки специальных текстов, машинного перевода и многоязыковой генерации текстов,
- овладеть лингвистической терминологией в области современных методов обработки текста на естественном языке,
- овладеть основами системы извлечения текстовой информации, извлечения знаний из больших массивов информации, работы с различными носителями информации, распределенными базами данных и знаний, глобальными компьютерными сетями,
- приобрести опыт работы с программным обеспечением для автоматического анализа текстов (морфологическими и синтаксическими анализаторами, конкордансами), классификации и тематического моделирования коллекций документов,
- овладеть методикой подготовки к переводу в части, касающейся компьютерной обработки информации,
- научиться оформлять текст перевода в разных типах компьютерных редакторов.

2 Место дисциплины в структуре образовательной программы

Дисциплина относится к дисциплинам (модулям) по выбору вариативной части блока 1 «Дисциплины (модули)»

Пререквизиты дисциплины: *Б.1.Б.10.1 Практический курс первого иностранного языка, Б.1.Б.10.2 Практический курс второго иностранного языка*

Постреквизиты дисциплины: *Отсутствуют*

3 Требования к результатам обучения по дисциплине

Процесс изучения дисциплины направлен на формирование следующих результатов обучения

Планируемые результаты обучения по дисциплине, характеризующие этапы формирования компетенций	Формируемые компетенции
Знать: основы современных технологий сбора, обработки и представления информации. Уметь: работать со стандартными компьютерными программами получения, обработки и управления информацией. Владеть: навыками получения, обработки и управления информацией.	ОПК-11 владением навыками работы с компьютером как средством получения, обработки и управления информацией
Знать: основные характеристики различных носителей информации, информационных баз, информационных процессов и глобальных компьютерных сетей. Уметь: работать с глобальными компьютерными сетями, распределёнными базами данных и знаний и вести расширенный поиск в поисковых системах глобальных сетей Интернет.	ОПК-12 способностью работать с различными носителями информации, распределенными базами данных и знаний, с глобальными компьютерными сетями

Планируемые результаты обучения по дисциплине, характеризующие этапы формирования компетенций	Формируемые компетенции
<u>Владеть:</u> методами, способами и средствами работы с носителями информации, распределенными базами данных и знаний, с глобальными компьютерными сетями для решения лингвистических задач, навыком работы с информацией в поисковых системах сети Интернет как глобальной базы данных и знаний; навыками работы с программными средствами общего и профессионального назначения.	
<u>Знать:</u> авторитетные электронные словари (в режиме он/оффлайн), справочно-информационные базы, лексические корпуса, поисковые системы; типы и принципы поисковых запросов. <u>Уметь:</u> использовать возможности, предоставляемые глобальными компьютерными сетями, для оптимального решения переводческих задач; пользоваться широким набором поисковых систем, электронных словарей и переводчиков, справочно-информационных баз, электронных ресурсов, параллельных корпусов; формулировать поисковые запросы и использовать поисковый инструментарий и справочный аппарат, составлять тезаурусы и глоссарии. <u>Владеть:</u> основными стратегиями поиска информации в разных типах источников; навыками работы с параллельными корпусами, автоматическими и автоматизированными системами перевода.	ПК-8 владением методикой подготовки к выполнению перевода, включая поиск информации в справочной, специальной литературе и компьютерных сетях
<u>Знать:</u> возможности различных видов компьютерных текстовых редакторов для оформления текста перевода; приемы редактирования и форматирования текста в различных текстовых редакторах. <u>Уметь:</u> использовать различные компьютерные текстовые редакторы для создания, просмотра, форматирования и редактирования разных типов текстов. <u>Владеть:</u> навыками нормативных требований к оформлению перевода в компьютерном текстовом редакторе: редактирования и форматирования текстов различной сложности; навыками внедрения в текст сторонних объектов (шрифтов, таблиц, формул, графиков, фотографий, изображений, диаграмм).	ПК-11 способностью оформлять текст перевода в компьютерном текстовом редакторе

4 Структура и содержание дисциплины

4.1 Структура дисциплины

Общая трудоемкость дисциплины составляет 3 зачетных единиц (108 академических часов).

Вид работы	Трудоемкость, академических часов	
	2 семестр	всего
Общая трудоёмкость	108	108
Контактная работа:	50,25	50,25
Практические занятия (ПЗ)	50	50
Промежуточная аттестация (зачет, экзамен)	0,25	0,25

Вид работы	Трудоемкость, академических часов	
	2 семестр	всего
Самостоятельная работа: - выполнение индивидуального творческого задания (ИТЗ); - самоподготовка (подготовка к практическим занятиям)	57,75	57,75
Вид итогового контроля (зачет, экзамен, дифференцированный зачет)	зачет	

Разделы дисциплины, изучаемые в 2 семестре

№ раздела	Наименование разделов	Количество часов				
		всего	аудиторная работа			внеауд. работа
			Л	ПЗ	ЛР	
1	Системы компьютерной (автоматической) обработки текста	14		6		8
2	Источники и носители информации	20		10		10
3	Типы лингвистических словарей. Онлайн словари	12		6		6
4	Системы аннотирования и реферирования	8		4		4
5	Машинный и автоматизированный перевод как система обработки текста	22		10		12
6	Оформление текста перевода в компьютерном текстовом редакторе	18		8		10
7	Параллельный корпус как инструмент переводчика	14		6		8
	Итого:	108		50		58
	Всего:	108		50		58

4.2 Содержание разделов дисциплины

1. Системы компьютерной (автоматической) обработки текста

Лингвистический автомат. Классификация систем КОТ по выполняемым функциям. Системы первого типа (программы Pragma, Lingvo, Белазар, Prompt, Socrat, Google Translate и др.). Системы генерации (синтеза) текстов по особому формальному описанию (программы SciGen, Textgen, VIRGO 4, Кобра 9.1). Трудности создания полноценно работающих с языком систем КОТ. Основные направления применения систем КОТ. Опознавание языка документа и нормализация входного текста. Подблок опознавания формата текста и его частей, а также определение их жанровой и тематической принадлежности. Моделирование языка с помощью конечного автомата. Формальные языки. Распознавание как поиск. Построение лексикона на основе конечного автомата.

2. Источники и носители информации

Типы носителей информации. Базы данных и знаний. Устройство глобальных компьютерных сетей. Поиск и обработка информации с помощью компьютера. Лингвистическая и техническая подготовка к выполнению перевода: составление глоссариев, тезаурусов, поиск справочной информации, работа с электронными словарями. Принципы работы терминологических баз.

3. Типы лингвистических словарей. Онлайн словари

Макроструктура словаря. Микроструктура словаря. Типология лингвистических словарей. Статистика речи и автоматический анализ текста. Лингвистический мониторинг функционирования языка. Компьютерное моделирование языка и речи.

4. Системы аннотирования и реферирования

Подходы и методы автоматического реферирования. Экстрактивные методы. Статистические методы. Логико-математические методы. Дистрибутивный метод. Метод содержательных аспектов. Метод текстовых связей. Методы с опорой на знания.

5. Машинный и автоматизированный перевод как система обработки текста

Устройство программ машинного и автоматизированного перевода. Подъязыки технической документации. Предредактирование. Постредактирование. Основные системы Translation Memory.

6. Оформление текста перевода в компьютерном текстовом редакторе

Типы текстовых редакторов. Microsoft Office Word. Ассистент слияния. Автотекст. Стили. Кадры. Шаблоны. Формы. Макрокоманды. Главный документ. Таблицы. Связь с другими документами.

7. Параллельный корпус как инструмент переводчика

Понятие корпусной лингвистики. Национальный корпус русского языка. Британский национальный корпус (British National Corpus, BNC). Лингвистические исследования на основе корпуса. Конкорданс как инструмент переводчика.

4.3 Практические занятия (семинары)

№ занятия	№ раздела	Тема	Кол-во часов
1-2	1	Системы КОТ	4
3	1	Конечные автоматы	2
4-5	2	Носители информации, базы данных и знаний, глобальные компьютерные сети	4
6-8	2	Подготовка к выполнению перевода	6
9-11	3	Работа с электронными словарями	6
12-13	4	Системы аннотирования и реферирования	4
14-15	5	Машинный перевод	4
16-18	5	Автоматизированный перевод	6
19-20	6	Основные параметры форматирования текста в компьютерном текстовом редакторе	4
21-22	6	Внедрение сторонних объектов	4
23-25	7	Параллельный корпус как инструмент переводчика	6
		Итого:	50

5 Учебно-методическое обеспечение дисциплины

5.1 Основная литература

1. Баймуратова, У.С. Электронный инструментальный переводчика: учебное пособие для студентов, обучающихся по программам высшего профессионального образования по направлению подготовки 035700.62 Лингвистика / У. С. Баймуратова. – Оренбург: Университет, 2013. – 120 с.

5.2 Дополнительная литература

1. Зубов, А.В. Информационные технологии в лингвистике / А.В. Зубов, И.И. Зубова. – М. : Академия, 2004. – 207 с. : ил. – (Высшее профессиональное образование). – Библиогр.: с. 192-204. – ISBN 5-7695-1531-7.

2. Овчинникова, И.Г. Компьютерное моделирование вербальной коммуникации: учеб.-метод. пособие / И. Г. Овчинникова, И.А. Углова. – М. : Флинта ; : Наука, 2009. – 136 с. – ISBN 978-5-9765-0729-6. – ISBN 978-5-02-034802-8.

3. Семенов, А.Л. Современные информационные технологии и перевод : учеб. пособие для вузов / А.Л. Семенов. – М. : Академия, 2008. – 224 с. – ISBN 978-5-7695-4459-0.

4. Хроленко, А.Т. Современные информационные технологии для гуманитария: практ. руководство / А.Т. Хроленко, А.В. Денисов. – М. : Флинта, 2008. – 128 с. – ISBN 978-5-9765-0023-5. – ISBN 978-5-02-034571-3.

5. Баранов, А.Н. Введение в прикладную лингвистику / А.Н. Баранов. – М.: ЛКИ, 2007. – 360 с. – ISBN 978-5-382-00231-6.

5.3 Периодические издания

Вестник компьютерных и информационных технологий : журнал. - Москва : Агентство "Роспечать", 2010-2016

5.4 Интернет-ресурсы

1. <https://www.multitran.ru/> – словарь для переводчиков «Мультитран»
2. <https://www.sdltrados.com/products/multiterm-desktop/> – инструмент управления терминологией SDL
3. <http://www.translate.ru> – онлайн-переводчик и словарь PROMT
4. <https://ru.smartcat.ai/> – система TM Smartcat
5. <http://translate.google.com> онлайн-переводчик Google Translate
6. <http://www.ruscorpora.ru> – Национальный корпус русского языка
7. <http://www.natcorp.ox.ac.uk/> – Британский национальный корпус

5.5 Программное обеспечение, профессиональные базы данных и информационные справочные системы современных информационных технологий

Лицензионное программное обеспечение

Операционная система Microsoft Windows

Пакет настольных приложений Microsoft Office (Word, Excel, PowerPoint, OneNote, Outlook, Publisher, Access)

Свободное программное обеспечение

1. Служебное и офисное ПО:

- Бесплатное средство просмотра файлов PDF Adobe Reader. Доступна бесплатно после принятия условий лицензионного соглашения на ПО Adobe. Разработчик: Adobe Systems. Режим доступа: <https://get.adobe.com/ru/reader/>.

- Свободный файловый архиватор 7-Zip. Предоставляется по лицензии GNU LGPL. Разработчик: Игорь Павлов. Режим доступа: <http://www.7-zip.org/>.

2. Электронные словари и переводчики:

- Свободная система автоматизированного перевода OmegaT. Предоставляется по лицензии GNU LGPL. Разработчики: Проект OmegaT поддерживается неофициальной международной группой добровольцев. Режим доступа: <http://www.omegat.org/>.

- Мультиплатформенная программа для проведения корпусных лингвистических исследований и управления данными AntConc. Доступна бесплатно после принятия условий лицензионного соглашения. Разработчик: Laurence Anthony (Center for English Language Education (CELESE), Faculty of Science and Engineering Waseda University). Режим доступа: <http://www.laurenceanthony.net/software.html>.

Профессиональные базы данных и информационные справочные системы современных информационных технологий:

1. SCOPUS [Электронный ресурс] : реферативная база данных / компания Elsevier. – Режим доступа: <https://www.scopus.com/>, в локальной сети ОГУ.

2. Springer [Электронный ресурс] : база данных научных книг, журналов, справочных материалов / компания Springer Customer Service Center GmbH . – Режим доступа : <https://link.springer.com/>, в локальной сети ОГУ.

6 Материально-техническое обеспечение дисциплины

Учебные аудитории для проведения занятий семинарского типа, для проведения групповых и индивидуальных консультаций, текущего контроля и промежуточной аттестации.

Аудитории оснащены комплектами ученической мебели, техническими средствами обучения, служащими для представления учебной информации большой аудитории.

Помещение для самостоятельной работы обучающихся оснащены компьютерной техникой, подключенной к сети "Интернет", и обеспечением доступа в электронную информационно-образовательную среду ОГУ.