

Минобрнауки России

Федеральное государственное бюджетное образовательное учреждение
высшего образования

«Оренбургский государственный университет»

Кафедра романской филологии и методики преподавания французского языка

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 (французский язык, второй иностранный язык)
(наименование направленности (профиля) образовательной программы)

Тип образовательной программы

Программа академического бакалавриата

Квалификация

Бакалавр

Форма обучения

Очная

Год набора 2018

Рабочая программа рассмотрена и утверждена на заседании кафедры

Кафедра романской филологии и методики преподавания французского языка

наименование кафедры

протокол № 6 от "22" 01 2018г.

Заведующий кафедрой

Кафедра романской филологии и методики преподавания французского языка

наименование кафедры



И.Ю. Моисеева

расшифровка подписи

Исполнители:

Профессор

должность



подпись

И.Ю. Моисеева

расшифровка подписи

должность

подпись

расшифровка подписи

СОГЛАСОВАНО:

Председатель методической комиссии по направлению подготовки

45.03.02 Лингвистика

код наименование

личная подпись

расшифровка подписи

И.А. Солодилова

Заведующий отделом комплектования научной библиотеки



личная подпись

Н.Н. Грицай

расшифровка подписи

Уполномоченный по качеству факультета



личная подпись

Т.В. Сапух

расшифровка подписи

№ регистрации _____

© Моисеева И.Ю., 2018

© ОГУ, 2018

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:

- владеть основами автоматического индексирования и реферирования текстов, предредактирования и постредактирования,
 - владеть лингвистической и статистической терминологией в области современных методов обработки текста на естественном языке,
 - уметь проводить статистическую обработку специальных текстов, машинный перевод и многоязыковую генерацию текстов,
 - владеть основами системы извлечения текстовой информации, извлечения знаний из больших массивов информации,
 - иметь навыки (приобрести опыт) работы с программным обеспечением для автоматического анализа текстов: морфологическими и синтаксическими анализаторами, конкордансами, классификации и тематического моделирования коллекций документов.
- ## 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 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. Владеть: методами, способами и средствами получения, хранения и переработки информации в области лингвистики; навыком работы с информацией в поисковых системах сети Интернет; навыками работы с программными средствами общего и профессионального назначения, основными навыками использования сети Интернет для поиска информации.	ОПК-12 способностью работать с различными носителями информации, распределёнными базами данных и знаний, с глобальными компьютерными сетями
Знать:	ПК-8 владением методикой

Планируемые результаты обучения по дисциплине, характеризующие этапы формирования компетенций	Формируемые компетенции
специальную профессиональную литературу для профессионально-ориентированного перевода, авторитетные электронные словари (в режиме он/оффлайн), справочно-информационные базы, лексические корпусы, поисковые системы; типы и принципы поисковых запросов. Уметь: использовать возможности, предоставляемые глобальными компьютерными сетями и справочной литературой, 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 инструментальный и справочный аппарат, составлять тезаурусы и глоссарии. Владеть: основными стратегиями поиска информации в разных типах источников; навыками работы с параллельными корпусами, автоматическими и автоматизированными системами перевода.	подготовки к выполнению перевода, включая поиск информации в справочной, специальной литературе и компьютерных сетях
Знать: возможности различных видов компьютерных текстовых редакторов для оформления текста перевода; приемы редактирования и форматирования текста в различных текстовых редакторах. Уметь: использовать различные компьютерные текстовые редакторы для создания, просмотра, форматирования и редактирования разных типов текстов. Владеть: 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
Самостоятельная работа: - выполнение индивидуального творческого задания (ИТЗ); - самоподготовка (проработка и повторение материала учебников и учебных пособий; - подготовка к практическим занятиям).	57,75	57,75
Вид итогового контроля (зачет, экзамен, дифференцированный зачет)	зачет	

Разделы дисциплины, изучаемые в 2 семестре

№	Наименование разделов	Количество часов
---	-----------------------	------------------

раздела		всего	аудиторная работа			внеауд. работа
			Л	ПЗ	ЛР	
1	Системы автоматической обработки текста и лингвистический автомат	8	-	4	-	4
2	Подъязыки в системах машинного перевода	8	-	4	-	4
3	Теория конечных автоматов	14	-	6	-	8
4	Системы аннотирования и реферирования	14	-	6	-	8
5	Компьютерная (автоматическая) обработка текстов	14	-	6	-	8
6	Типы лингвистических словарей. Онлайн словари и переводчики	18	-	8	-	10
7	Оформление текста перевода в компьютерном текстовом редакторе. Работа с текстовым редактором Microsoft Office	20	-	10	-	10
8	Параллельный корпус как инструмент переводчика	12	-	6	-	6
	Итого:	108	-	50	-	58
	Всего:	108	-	50	-	58

4.2 Содержание разделов дисциплины

1. Системы автоматической обработки текста и лингвистический автомат

Лингвистический автомат. Человеко-машинная синергетика и самоорганизация в системах АОТ. Уровневое построение систем АОТ и ЛА. Опознавание языка документа и нормализация входного текста. Подблок опознавания формата текста и его частей, а также определение их жанровой и тематической принадлежности. Средства обеспечения живучести ЛА.

2. Подъязыки в системах машинного перевода

Подъязыки технической документации. Предредактирование. Постредактирование.

3. Теория конечных автоматов.

Моделирование языка с помощью конечного автомата. Формальные языки. Недетерминированные конечные автоматы. Распознавание как поиск. Построение лексикона на основе конечного автомата.

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. Компьютерная (автоматическая) обработка текстов.

Классификация систем КОТ по выполняемым функциям. Системы первого типа (программы Pragma, Lingvo, Белазар, Promt, 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.

6. Типы лингвистических словарей. Онлайн словари и переводчики

Макроструктура словаря. Микроструктура словаря. Типология лингвистических словарей. Компьютерная обработка текста: описание моделей и направлений разработок. Статистика речи и автоматический анализ текста. Лингвистический мониторинг функционирования языка. Компьютерное моделирование языка и речи.

7. Оформление текста перевода в компьютерном текстовом редакторе. Работа с текстовым редактором Microsoft Office

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

8. 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
-----------	-----------	------	--------------

№ занятия	№ раздела	Тема	Кол-во часов
1-2	1	Системы автоматической обработки текста и лингвистический автомат	4
3-4	2	Подъязыки в системах машинного перевода	4
5-7	3	Теория конечных автоматов	6
8-10	4	Системы аннотирования и реферирования	6
11-13	5	Компьютерная (автоматическая) обработка текстов	6
14-17	6	Типы лингвистических словарей. Онлайн словари и переводчики	8
18-22	7	Оформление текста перевода в компьютерном текстовом редакторе. Работа с текстовым редактором Microsoft Office	10
23-25	8	Параллельный корпус как инструмент переводчика	6
		Итого:	50

5 Учебно-методическое обеспечение дисциплины

5.1. Основная литература

Мишова, В.В. Мультимедийные технологии : практикум / В.В. Мишова ; Министерство культуры Российской Федерации, Кемеровский государственный институт культуры, Институт информационных и библиотечных технологий. - Кемерово : Кемеровский государственный институт культуры, 2017. - 80 с. : ил. - Библиогр.: с. 78. - ISBN 978-5-8154-0374-1 ; То же [Электронный ресурс]. - URL: <http://biblioclub.ru/index.php?page=book&id=472682>

Нужнов, Е.В. Мультимедиа технологии : учебное пособие / Е.В. Нужнов ; Министерство образования и науки РФ, Южный федеральный университет. - 2-е изд., перераб. и доп. - Таганрог : Издательство Южного федерального университета, 2016. - Ч. 2. Виртуальная реальность, создание мультимедиа продуктов, применение мультимедиа технологий в профессиональной деятельности. - 180 с. : ил. - Библиогр. в кн. - ISBN 978-5-9275-2171-5 ; То же [Электронный ресурс]. - URL: <http://biblioclub.ru/index.php?page=book&id=493255>

5.2 Дополнительная литература

1 Колокольникова, А.И. Информатика : учебное пособие / А.И. Колокольникова, Е.В. Прокопенко, Л.С. Таганов. - Москва : Директ-Медиа, 2013. - 115 с. - Библиогр. в кн. - ISBN 978-5-4458-2864-8 ; То же [Электронный ресурс]. - URL: <http://biblioclub.ru/index.php?page=book&id=210626>

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

3. Онтологии и тезаурусы: модели, инструменты, приложения : учебное пособие / Б.В. Добров, В.В. Иванов, Н.В. Лукашевич, В.Д. Соловьев. - Москва : Интернет-Университет Информационных Технологий, 2009. - 173 с. : ил.,табл., схем. - (Основы информационных технологий). - ISBN 978-5-9963-0007-5 ; То же [Электронный ресурс]. - URL: <http://biblioclub.ru/index.php?page=book&id=233056>

4. 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.

5. Формализация исследовательских процедур анализа семантики языковых единиц : коллективная монография / М.В. Каменский, Т.Н. Ломтева, Н.С. Кабылкина и др. ; под общ. ред. М.В. Каменского ; Министерство образования и науки РФ, Федеральное государственное автономное образовательное учреждение высшего образования «Северо-Кавказский федеральный университет». - Ставрополь : СКФУ, 2016. - 170 с. : ил. - Библиогр.: с.166-167 - ISBN 978-5-9296-0838-4 ; То же [Электронный ресурс]. - URL: <http://biblioclub.ru/index.php?page=book&id=466913>

5.3 Периодические издания

Вестник компьютерных и информационных технологий : журнал. - Москва : Агентство "Роспечать", 2018.

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. <http://translate.google.com> онлайн-переводчик Google Translate
5. <http://www.ruscorpora.ru> – Национальный корпус русского языка

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, PowerPoint)

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/>.

Профессиональные базы данных и информационные справочные системы современных информационных технологий:

1. <http://inion.ru/> - Крупнейший в России комплекс библиографических баз данных по гуманитарным и социальным наукам, который ведется с 1980 года. В БД включаются аннотированные описания книг и статей из журналов и сборников на 140 языках мира, поступающих в библиотеку ИНИОН. На сервере ИНИОН предоставляется свободный доступ к нескольким сводным каталогам, отражающим поступление литературы за определенные периоды времени (1993-1995, 1996-1998, 1999-2000 годы).

2. <https://www.scopus.com/search/form.uri?display=basic> (доступ открыт из сети университета) - Крупнейшая база аннотаций и цитируемости рецензируемой научной литературы со встроенными инструментами мониторинга, анализа и визуализации научно-исследовательских данных.

3. <http://apps.webofknowledge.com/> - (доступ открыт из сети университета) Пакет наукометрических ресурсов компании Thomson Reuters. Цитатные базы данных Web of Science включают списки всех библиографических ссылок, встречающихся в научных публикациях: статьях, материалах конференций, семинаров, симпозиумов.

4. Гарант [Электронный ресурс] : справочно-правовая система / НПП Гарант- Сервис. – Электрон. дан. - Москва, [1990–2018]. – Режим доступа \\fileserver1\GarantClient\garant.exe

5. Консультант Плюс [Электронный ресурс] : справочно-правовая система / Компания Консультант Плюс. – Электрон. дан. – Москва, [1992–2018]. – Режим доступа : в локальной сети ОГУ \\fileserver1\CONSULT\cons.exe

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.

К рабочей программе прилагаются:

- Фонд оценочных средств для проведения промежуточной аттестации обучающихся по дисциплине;
- Методические указания для обучающихся по освоению дисциплины.