

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ
Федеральное государственное автономное образовательное учреждение высшего образования
"Казанский (Приволжский) федеральный университет"
Институт филологии и межкультурной коммуникации
Высшая школа русской филологии и культуры им. Льва Толстого



УТВЕРЖДАЮ

Проректор по образовательной деятельности КФУ

_____ Турилова Е.А.

"__" _____ 20__ г.

Программа дисциплины
Методы обработки естественного языка

Направление подготовки: 45.04.01 - Филология

Профиль подготовки: Прикладная лингвистика (с применением электронного обучения и дистанционных образовательных технологий)

Квалификация выпускника: магистр

Форма обучения: заочное

Язык обучения: русский

Год начала обучения по образовательной программе: 2024

Содержание

1. Перечень планируемых результатов обучения по дисциплине (модулю), соотнесенных с планируемыми результатами освоения ОПОП ВО
2. Место дисциплины (модуля) в структуре ОПОП ВО
3. Объем дисциплины (модуля) в зачетных единицах с указанием количества часов, выделенных на контактную работу обучающихся с преподавателем (по видам учебных занятий) и на самостоятельную работу обучающихся
4. Содержание дисциплины (модуля), структурированное по темам (разделам) с указанием отведенного на них количества академических часов и видов учебных занятий
 - 4.1. Структура и тематический план контактной и самостоятельной работы по дисциплине (модулю)
 - 4.2. Содержание дисциплины (модуля)
5. Перечень учебно-методического обеспечения для самостоятельной работы обучающихся по дисциплине (модулю)
6. Фонд оценочных средств по дисциплине (модулю)
7. Перечень литературы, необходимой для освоения дисциплины (модуля)
8. Перечень ресурсов информационно-телекоммуникационной сети "Интернет", необходимых для освоения дисциплины (модуля)
9. Методические указания для обучающихся по освоению дисциплины (модуля)
10. Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине (модулю), включая перечень программного обеспечения и информационных справочных систем (при необходимости)
11. Описание материально-технической базы, необходимой для осуществления образовательного процесса по дисциплине (модулю)
12. Средства адаптации преподавания дисциплины (модуля) к потребностям обучающихся инвалидов и лиц с ограниченными возможностями здоровья
13. Приложение №1. Фонд оценочных средств
14. Приложение №2. Перечень литературы, необходимой для освоения дисциплины (модуля)
15. Приложение №3. Перечень информационных технологий, используемых для освоения дисциплины (модуля), включая перечень программного обеспечения и информационных справочных систем

Программу дисциплины разработал(а)(и): доцент, к.н. Вольская Ю.А. (кафедра прикладной и экспериментальной лингвистики, Высшая школа русской филологии и культуры им Льва Толстого), Kovaleva95julia@mail.ru

1. Перечень планируемых результатов обучения по дисциплине (модулю), соотнесенных с планируемыми результатами освоения ОПОП ВО

Обучающийся, освоивший дисциплину (модуль), должен обладать следующими компетенциями:

Шифр компетенции	Расшифровка приобретаемой компетенции
ОПК-2	Способен использовать в профессиональной деятельности, в том числе педагогической, знания современной научной парадигмы в области филологии и динамики ее развития, системы методологических принципов и методических приемов филологического исследования
УК-5	Способен анализировать и учитывать разнообразие культур в процессе межкультурного взаимодействия

Обучающийся, освоивший дисциплину (модуль):

Должен знать:

- современную теоретическую концепцию культуры речи, все основные нормы русского литературного языка в связи с использованием методов обработки естественного языка
- методы научного исследования в области компьютерной лингвистики и обработки естественного языка.

Должен уметь:

- эффективно находить, воспринимать и использовать информацию на русском языке, полученную из печатных и электронных источников для решения задач обработки естественного языка
- представлять результаты собственного исследования в виде письменных жанров научной коммуникации, интегрировать оптимальные зарубежные и отечественные информационно-коммуникационные технологии в процессе сбора, интерпретации, систематизации и анализа научной литературы в области обработки естественного языка.

Должен владеть:

- способами критической оценки эффективности различных методов и технологий обработки естественного языка
- навыками виртуального представления результатов собственного исследования в области обработки естественного языка и навыками публичного выступления по теме собственного исследования

2. Место дисциплины (модуля) в структуре ОПОП ВО

Данная дисциплина (модуль) включена в раздел "Б1.О.04.02 Дисциплины (модули)" основной профессиональной образовательной программы 45.04.01 "Филология (Прикладная лингвистика (с применением электронного обучения и дистанционных образовательных технологий))" и относится к обязательной части ОПОП ВО.

Осваивается на 3 курсе в 5 семестре.

3. Объем дисциплины (модуля) в зачетных единицах с указанием количества часов, выделенных на контактную работу обучающихся с преподавателем (по видам учебных занятий) и на самостоятельную работу обучающихся

Общая трудоемкость дисциплины составляет 4 зачетных(ые) единиц(ы) на 144 часа(ов).

Контактная работа - 25 часа(ов), в том числе лекции - 8 часа(ов), практические занятия - 16 часа(ов), лабораторные работы - 0 часа(ов), контроль самостоятельной работы - 1 часа(ов).

Самостоятельная работа - 115 часа(ов).

Контроль (зачёт / экзамен) - 4 часа(ов).

Форма промежуточного контроля дисциплины: зачет в 5 семестре.

4. Содержание дисциплины (модуля), структурированное по темам (разделам) с указанием отведенного на них количества академических часов и видов учебных занятий

4.1 Структура и тематический план контактной и самостоятельной работы по дисциплине (модулю)

N	Разделы дисциплины / модуля	Се- местр	Виды и часы контактной работы, их трудоемкость (в часах)						Само- стоя- тель- ная ра- бота
			Лекции, всего	Лекции в эл. форме	Практи- ческие занятия, всего	Практи- ческие в эл. форме	Лабора- торные работы, всего	Лабора- торные в эл. форме	
1.	Тема 1. История развития NLP. Общие этапы и модули обработки текстов	5	2	0	6	0	0	0	38
2.	Тема 2. Морфологический анализ текстов. Парсинг	5	4	0	6	0	0	0	39
3.	Тема 3. Семантическая и информационная структуры при анализе текстов. Семантические сети	5	2	0	4	0	0	0	38
	Итого		8	0	16	0	0	0	115

4.2 Содержание дисциплины (модуля)

Тема 1. История развития NLP. Общие этапы и модули обработки текстов

Компьютерная лингвистика и обработка естественного языка. Основные этапы развития обработки естественного языка. Задачи обработки естественного языка. Актуальное состояние NLP, нерешенные задачи в области обработки русского языка. Извлечение сущностей. Диалоговые системы. Ответы на вопросы. Нейронные сети в NLP. Основные этапы обработки текстов. Базовая терминология. Составление словаря: токенизация.

Тема 2. Морфологический анализ текстов. Парсинг

Обзор модулей морфологического анализа. Морфологические модули AOT. Морфологический анализатор Rymorphy2. Словарь проекта OpenCorpora. Анализатор Mystem. Статистическая частеречная разметка. Триграммная скрытая Марковская модель. Методы автоматического снятия омонимии. Парсинг. Грамматика зависимостей. Грамматика составляющих. Комбинированный подход к парсингу. Статистический парсинг. Обзор модулей. Лингвистический процессор ЭТАП. DictaScore и AOT. Link Grammar Parser.

Тема 3. Семантическая и информационная структуры при анализе текстов. Семантические сети

Модели представления знаний в компьютерной семантике. Сетевые представления. Концептуальные графы. Теория графов. Фреймовые сети. Формальные онтологии. Методы автоматического построения формальных онтологий. Обзор современных онтологий. Тезаурусы типа WordNet. Анализ текста в парадигме когнитивных исследований. Анализ текста в парадигмах автоматического понимания текста. Дискурсивный анализ в NLP. Векторное представление слов. Формальная семантика.

5. Перечень учебно-методического обеспечения для самостоятельной работы обучающихся по дисциплине (модулю)

Самостоятельная работа обучающихся выполняется по заданию и при методическом руководстве преподавателя, но без его непосредственного участия. Самостоятельная работа подразделяется на самостоятельную работу на аудиторных занятиях и на внеаудиторную самостоятельную работу. Самостоятельная работа обучающихся включает как полностью самостоятельное освоение отдельных тем (разделов) дисциплины, так и проработку тем (разделов), осваиваемых во время аудиторной работы. Во время самостоятельной работы обучающиеся читают и конспектируют учебную, научную и справочную литературу, выполняют задания, направленные на закрепление знаний и отработку умений и навыков, готовятся к текущему и промежуточному контролю по дисциплине.

Организация самостоятельной работы обучающихся регламентируется нормативными документами, учебно-методической литературой и электронными образовательными ресурсами, включая:

Порядок организации и осуществления образовательной деятельности по образовательным программам высшего образования - программам бакалавриата, программам специалитета, программам магистратуры (утвержден приказом Министерства науки и высшего образования Российской Федерации от 6 апреля 2021 года №245)

Письмо Министерства образования Российской Федерации №14-55-99бин/15 от 27 ноября 2002 г. "Об активизации самостоятельной работы студентов высших учебных заведений"

Устав федерального государственного автономного образовательного учреждения "Казанский (Приволжский) федеральный университет"

Правила внутреннего распорядка федерального государственного автономного образовательного учреждения высшего профессионального образования "Казанский (Приволжский) федеральный университет"

Локальные нормативные акты Казанского (Приволжского) федерального университета

6. Фонд оценочных средств по дисциплине (модулю)

Фонд оценочных средств по дисциплине (модулю) включает оценочные материалы, направленные на проверку освоения компетенций, в том числе знаний, умений и навыков. Фонд оценочных средств включает оценочные средства текущего контроля и оценочные средства промежуточной аттестации.

В фонде оценочных средств содержится следующая информация:

- соответствие компетенций планируемым результатам обучения по дисциплине (модулю);
- критерии оценивания сформированности компетенций;
- механизм формирования оценки по дисциплине (модулю);
- описание порядка применения и процедуры оценивания для каждого оценочного средства;
- критерии оценивания для каждого оценочного средства;
- содержание оценочных средств, включая требования, предъявляемые к действиям обучающихся, демонстрируемым результатам, задания различных типов.

Фонд оценочных средств по дисциплине находится в Приложении 1 к программе дисциплины (модулю).

7. Перечень литературы, необходимой для освоения дисциплины (модуля)

Освоение дисциплины (модуля) предполагает изучение основной и дополнительной учебной литературы. Литература может быть доступна обучающимся в одном из двух вариантов (либо в обоих из них):

- в электронном виде - через электронные библиотечные системы на основании заключенных КФУ договоров с правообладателями;
- в печатном виде - в Научной библиотеке им. Н.И. Лобачевского. Обучающиеся получают учебную литературу на абонементе по читательским билетам в соответствии с правилами пользования Научной библиотекой.

Электронные издания доступны дистанционно из любой точки при введении обучающимся своего логина и пароля от личного кабинета в системе "Электронный университет". При использовании печатных изданий библиотечный фонд должен быть укомплектован ими из расчета не менее 0,5 экземпляра (для обучающихся по ФГОС 3++ - не менее 0,25 экземпляра) каждого из изданий основной литературы и не менее 0,25 экземпляра дополнительной литературы на каждого обучающегося из числа лиц, одновременно осваивающих данную дисциплину.

Перечень основной и дополнительной учебной литературы, необходимой для освоения дисциплины (модуля), находится в Приложении 2 к рабочей программе дисциплины. Он подлежит обновлению при изменении условий договоров КФУ с правообладателями электронных изданий и при изменении комплектования фондов Научной библиотеки КФУ.

8. Перечень ресурсов информационно-телекоммуникационной сети "Интернет", необходимых для освоения дисциплины (модуля)

MyStem - <https://yandex.ru/dev/mystem/>

RuLingva - <https://rulex.kpfu.ru/>

Автоматическая обработка текстов (АОТ) - <http://www.aot.ru/onlinedemo.html>

9. Методические указания для обучающихся по освоению дисциплины (модуля)

Вид работ	Методические рекомендации
лекции	В процессе посещения лекционных занятий необходимо вести конспектирование учебного материала, обращать внимание на формулировки терминов, понятий, основных категорий компьютерной лингвистики. Необходимо помнить, что посещение лекций является обязательным, и, в случае пропуска занятия, обучающийся должен изучить его содержание самостоятельно.
практические занятия	При подготовке к практическим занятиям по данной дисциплине желательно опираться на учебники, представленные в списке литературы. Особое внимание при изучении дисциплины студентам следует обратить на вопросы, связанные с частными методами автоматической обработки текстов, а также на темы в области актуальных направлений NLP.

Вид работ	Методические рекомендации
самостоятельная работа	Самостоятельная работа рассматривается как важнейшее звено в освоении дисциплины. В рамках самостоятельной работы рекомендуется чтение и конспектирование рекомендованной по курсу литературы с последующим составлением плана или умственной схемы по той или иной теме. Преподаватель может помочь в поиске литературы по теме. Контроль выполнения самостоятельной работы студентов (в виде устного или письменного опроса) обязателен.
зачет	Зачет сдается по билетам. Для успешной сдачи зачета студенту необходимо свободно владеть терминологией компьютерной и корпусной лингвистики, иметь представление о проблемах прикладной лингвистики, ориентироваться в этапах становления различных направлений компьютерной лингвистики, иметь представление о проблемах корпусной лингвистики, основных корпусах русского, английского, немецкого, чешского языков, владеть терминологией данной дисциплины и др.

10. Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине (модулю), включая перечень программного обеспечения и информационных справочных систем (при необходимости)

Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине (модулю), включая перечень программного обеспечения и информационных справочных систем, представлен в Приложении 3 к рабочей программе дисциплины (модуля).

11. Описание материально-технической базы, необходимой для осуществления образовательного процесса по дисциплине (модулю)

Материально-техническое обеспечение образовательного процесса по дисциплине (модулю) включает в себя следующие компоненты:

Помещения для самостоятельной работы обучающихся, укомплектованные специализированной мебелью (столы и стулья) и оснащенные компьютерной техникой с возможностью подключения к сети "Интернет" и обеспечением доступа в электронную информационно-образовательную среду КФУ.

Учебные аудитории для контактной работы с преподавателем, укомплектованные специализированной мебелью (столы и стулья).

Компьютер и принтер для распечатки раздаточных материалов.

12. Средства адаптации преподавания дисциплины к потребностям обучающихся инвалидов и лиц с ограниченными возможностями здоровья

При необходимости в образовательном процессе применяются следующие методы и технологии, облегчающие восприятие информации обучающимися инвалидами и лицами с ограниченными возможностями здоровья:

- создание текстовой версии любого нетекстового контента для его возможного преобразования в альтернативные формы, удобные для различных пользователей;
- создание контента, который можно представить в различных видах без потери данных или структуры, предусмотреть возможность масштабирования текста и изображений без потери качества, предусмотреть доступность управления контентом с клавиатуры;
- создание возможностей для обучающихся воспринимать одну и ту же информацию из разных источников - например, так, чтобы лица с нарушениями слуха получали информацию визуально, с нарушениями зрения - аудиально;
- применение программных средств, обеспечивающих возможность освоения навыков и умений, формируемых дисциплиной, за счёт альтернативных способов, в том числе виртуальных лабораторий и симуляционных технологий;
- применение дистанционных образовательных технологий для передачи информации, организации различных форм интерактивной контактной работы обучающегося с преподавателем, в том числе вебинаров, которые могут быть использованы для проведения виртуальных лекций с возможностью взаимодействия всех участников дистанционного обучения, проведения семинаров, выступления с докладами и защиты выполненных работ, проведения тренингов, организации коллективной работы;
- применение дистанционных образовательных технологий для организации форм текущего и промежуточного контроля;
- увеличение продолжительности сдачи обучающимся инвалидом или лицом с ограниченными возможностями здоровья форм промежуточной аттестации по отношению к установленной продолжительности их сдачи;
- продолжительности сдачи зачёта или экзамена, проводимого в письменной форме, - не более чем на 90 минут;
- продолжительности подготовки обучающегося к ответу на зачёте или экзамене, проводимом в устной форме, - не более чем на 20 минут;
- продолжительности выступления обучающегося при защите курсовой работы - не более чем на 15 минут.

Программа составлена в соответствии с требованиями ФГОС ВО и учебным планом по направлению 45.04.01 "Филология" и магистерской программе "Прикладная лингвистика (с применением электронного обучения и дистанционных образовательных технологий)".

*Приложение 2
к рабочей программе дисциплины (модуля)
Б1.О.04.02 Методы обработки естественного языка*

Перечень литературы, необходимой для освоения дисциплины (модуля)

Направление подготовки: 45.04.01 - Филология

Профиль подготовки: Прикладная лингвистика (с применением электронного обучения и дистанционных образовательных технологий)

Квалификация выпускника: магистр

Форма обучения: заочное

Язык обучения: русский

Год начала обучения по образовательной программе: 2024

Основная литература:

1. Корягин С. В., Русаков А. М. Методы анализа естественно-языковых текстов: Учебно-методическое пособие. Москва : РТУ МИРЭА, 2023. - 86 с. - ISBN 978-5-7339-1737-5- Текст : электронный. URL: <https://reader.lanbook.com/book/331649#2> - Режим доступа: по подписке.
2. Баланов А.Н. Машинное обучение и искусственный интеллект: Учебное пособие для вузов. М.: Издательство 'Лань' 2024. - 172 страниц - ISBN: 978-5-507-49194-0 - Текст : электронный. - URL: <https://reader.lanbook.com/book/414920#3> - Режим доступа: по подписке.
3. Надеина, Т. М. Основы прикладной и математической лингвистики : учебное пособие/Т. М. Надеина. - Москва : Норма : ИНФРА-М, 2020. - 152 с. - ISBN 978-5-00156-072-2. - Текст : электронный. - URL: <https://znanium.com/catalog/product/1092460> (дата обращения: 28.04.2020). - Режим доступа: по подписке.

Дополнительная литература:

1. Щипицина, Л. Ю. Информационные технологии в лингвистике : учебное пособие/Л. Ю. Щипицина. - Москва : ФЛИНТА, 2013. - 128 с. - ISBN 978-5-9765-1431-7. - Текст : электронный. - URL: <https://znanium.com/catalog/product/462989> (дата обращения: 28.04.2020). - Режим доступа: по подписке.
2. Гребенщикова, А. В. Основы квантитативной лингвистики и новых информационных технологий : учебное пособие/А. В. Гребенщикова. - 3-е изд., стер. - Москва : ФЛИНТА, 2018. - 152 с. - ISBN 978-5-9765-2137-7. - Текст : электронный. - URL: <https://znanium.com/catalog/product/1142454> (дата обращения: 28.04.2020). - Режим доступа: по подписке.

*Приложение 3
к рабочей программе дисциплины (модуля)
Б1.О.04.02 Методы обработки естественного языка*

Перечень информационных технологий, используемых для освоения дисциплины (модуля), включая перечень программного обеспечения и информационных справочных систем

Направление подготовки: 45.04.01 - Филология

Профиль подготовки: Прикладная лингвистика (с применением электронного обучения и дистанционных образовательных технологий)

Квалификация выпускника: магистр

Форма обучения: заочное

Язык обучения: русский

Год начала обучения по образовательной программе: 2024

Освоение дисциплины (модуля) предполагает использование следующего программного обеспечения и информационно-справочных систем:

Операционная система Microsoft Windows 7 Профессиональная или Windows XP (Volume License)

Пакет офисного программного обеспечения Microsoft Office 365 или Microsoft Office Professional plus 2010

Браузер Mozilla Firefox

Браузер Google Chrome

Adobe Reader XI или Adobe Acrobat Reader DC

Kaspersky Endpoint Security для Windows

Учебно-методическая литература для данной дисциплины имеется в наличии в электронно-библиотечной системе "ZNANIUM.COM", доступ к которой предоставлен обучающимся. ЭБС "ZNANIUM.COM" содержит произведения крупнейших российских учёных, руководителей государственных органов, преподавателей ведущих вузов страны, высококвалифицированных специалистов в различных сферах бизнеса. Фонд библиотеки сформирован с учетом всех изменений образовательных стандартов и включает учебники, учебные пособия, учебно-методические комплексы, монографии, авторефераты, диссертации, энциклопедии, словари и справочники, законодательно-нормативные документы, специальные периодические издания и издания, выпускаемые издательствами вузов. В настоящее время ЭБС ZNANIUM.COM соответствует всем требованиям федеральных государственных образовательных стандартов высшего образования (ФГОС ВО) нового поколения.

Учебно-методическая литература для данной дисциплины имеется в наличии в электронно-библиотечной системе Издательства "Лань", доступ к которой предоставлен обучающимся. ЭБС Издательства "Лань" включает в себя электронные версии книг издательства "Лань" и других ведущих издательств учебной литературы, а также электронные версии периодических изданий по естественным, техническим и гуманитарным наукам. ЭБС Издательства "Лань" обеспечивает доступ к научной, учебной литературе и научным периодическим изданиям по максимальному количеству профильных направлений с соблюдением всех авторских и смежных прав.

Учебно-методическая литература для данной дисциплины имеется в наличии в электронно-библиотечной системе "Консультант студента", доступ к которой предоставлен обучающимся. Многопрофильный образовательный ресурс "Консультант студента" является электронной библиотечной системой (ЭБС), предоставляющей доступ через сеть Интернет к учебной литературе и дополнительным материалам, приобретенным на основании прямых договоров с правообладателями. Полностью соответствует требованиям федеральных государственных образовательных стандартов высшего образования к комплектованию библиотек, в том числе электронных, в части формирования фондов основной и дополнительной литературы.