

Программа разработана экспертами
Федерального учебно-методического объединения
высшего образования по укрупненной группе
специальностей и направлений подготовки
45.00.00 Языкознание и литературоведение

Утверждена на заседании ФУМО
25 мая 2021 года

Примерная программа учебной дисциплины

**ТЕХНОЛОГИИ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА В
ГУМАНИТАРНЫХ ИССЛЕДОВАНИЯХ**

Уровень высшего образования:

БАКАЛАВРИАТ

Направление подготовки:

45.03.01 «ФИЛОЛОГИЯ»

45.03.02 «ЛИНГВИСТИКА»

Раздел 1. Характеристики учебных занятий

1.1. Цели и задачи учебных занятий

Целью данного курса является изучение технологий искусственного интеллекта в целях обучения применению таких технологий в различных сферах деятельности. В рамках курса предусматривается ознакомление учащихся с основными типами лингвистических информационных ресурсов и с инструментами, которые могут быть полезны при решении базовых филологических и лингвистических прикладных задач. Результатом занятий должно стать приобретение студентами навыков работы с лингвистическими базами данных, корпусами текстов и электронными словарями, а также получение представлений о возможном применении данных средств и ресурсов при проведении исследований.

1.2 Место дисциплины (модуля) в структуре образовательной программы, связь с другими дисциплинами (модулями) программы

Относится к базовой части ОПОП ВО.

1.3 Требования подготовленности обучающегося к освоению содержания учебных занятий (пререквизиты)

Устанавливаются образовательной организацией.

1.4 Перечень результатов обучения

По окончании курса студент должен уметь работать с технологиями искусственного интеллекта, филологическими и лингвистическими ресурсами, знать различные типы электронных лингвистических ресурсов, уметь сравнивать различные электронные ресурсы и выбирать из них наиболее подходящие для решения конкретных исследовательских задач, иметь навыки работы с электронными словарями, корпусами текстов, лингвистическими базами данных.

Набор компетенций, соотнесенных с результатами обучения определяется образовательной организацией.

1.5 Перечень рекомендуемых образовательных технологий

В преподавании дисциплины «Технологии искусственного интеллекта в гуманитарных исследованиях» используются разнообразные образовательные технологии как традиционного, так и инновационного характера, учитывающие смешанный, теоретико- и практикоориентированный характер дисциплины:

- лекции;
- практические занятия;
- дискуссии;
- выступления с докладами и сообщениями;
- аудиторные контрольные работы;
- внеаудиторные контрольные работы;
- тестирование.

Степень необходимости образовательной среды и ее выбор определяется образовательной организацией. Формы текущей аттестации определяются образовательной организацией.

1.6 Объем дисциплины (модуля) в зачетных единицах

2 з.е.

Раздел 2. Организация, структура и содержание учебных занятий

2.1 Организация учебных занятий

Предусмотрены учебные занятия с использованием дистанционных технологий.

2.2 Краткая аннотация содержания дисциплины (модуля)

Наименование темы (раздела, части)	Вид учебных занятий	Кол-во часов
1. Искусственный интеллект в гуманитарной сфере. Проблемы искусственного интеллекта и типы решаемых задач. Введение в цифровую филологию. Направления цифровой филологии.	Лекции	2
2. Филологические и лингвистические ресурсы и программы. Электронные лингвистические ресурсы. Цифровые библиотеки. Информационные системы в филологических задачах. Задачи информационного поиска с точки зрения филолога.	Лекции (2) Практические занятия (4)	6
3. Национальные и иные корпуса. Национальный корпус русского языка и другие русскоязычные корпуса. British National Corpus как первый пример национального корпуса. Корпусы иных языков. Обработка корпусных данных. Структура и назначение параллельных корпусов в решении задач ИИ. Примеры исследований с использованием корпусных данных.	Лекции (2) Практические занятия (10)	12
4. Google books Ngram Viewer и поиск словосочетаний. Google books Ngram Viewer. Сервисы исследования биграмм и n-грамм.	Лекции (2) Практические занятия (2)	4

<p>5. Компьютерная лексикография и системы представления лексических знаний. Электронные словари и их особенности. Электронные словари, доступные в сети. Частотные словари. Иноязычные словари. Справочные системы для изучающих язык. Лексическая информация в системах ИИ.</p>	<p>Лекции (2) Практические занятия (4)</p>	<p>6</p>
<p>6. Инструменты искусственного интеллекта. Прикладные аспекты ИИ. Лингвистические базы данных и экспертные системы. Подходы к классификации: обучение с учителем и обучение без учителя. Самообучающиеся системы: нейронные сети.</p>	<p>Лекции (2) Практические занятия (4)</p>	<p>6</p>
<p>ИТОГО:</p>		<p>36</p>

45.00.00

ЛИТЕРАТУРОВЕДЕНИЕ

Раздел 3. Обеспечение учебных занятий

3.1 Методические указания по освоению дисциплины

Преподавание дисциплины осуществляется в форме лекционных и практических занятий, которые сопровождаются демонстрацией наглядного материала – Интернет-ресурсов и презентаций о них. Во время занятий обучающиеся выполняют практические задания. Для закрепления пройденного материала предлагаются домашние задания по каждой из тем. Успешное овладение содержанием дисциплины «Технологии искусственного интеллекта в гуманитарных исследованиях» предполагает работу обучающихся в группах в аудитории, а также их самостоятельную работу.

Дополнительные методические указания устанавливаются образовательной организацией.

3.2 Примерный перечень учебно-методического обеспечения самостоятельной работы обучающихся по дисциплине (модулю), в том числе примерный перечень учебной литературы и ресурсов информационно-телекоммуникационной сети «Интернет», необходимых для освоения дисциплины (модуля)

Самостоятельная работа студентов должна включать усвоение теоретического материала, подготовку к практическим занятиям, выполнение творческих заданий, работу с электронным учебно-методическим комплексом, подготовку к текущему контролю знаний, к промежуточной аттестации (зачету).

Список рекомендованной литературы

- Беляева Л.Н. 2014. Лингвистические ресурсы информационной образовательной среды: состав, структура и функции. *Известия Российского государственного педагогического университета им. А.И. Герцена*. № 171. 47-52.
- Беляева Л.Н. 2015. Сетевые лингвистические ресурсы в подготовке филолога. *Вестник Пермского национального исследовательского политехнического университета. Проблемы языкознания и педагогики*. № 4 (14). 36-48.
- Дейт К. 2005. Введение в системы баз данных. М., СПб, Киев.
- Ляшевская О. Н., Шаров С. А. 2009. Частотный словарь русской лексики (на материалах Национального корпуса русского языка).
- Марчук Ю.Н. 2005. Компьютерная лингвистика. М.: Восток/Запад.
- Национальный корпус русского языка: 2003–2005. Результаты и перспективы. Сборник статей. М., 2005.
- Национальный корпус русского языка: 2006–2008. Новые результаты и перспективы. Сборник статей. СПб, 2009.
- Плунгян В. Почему современная лингвистика должна быть лингвистикой корпусов. URL: <http://www.polit.ru/lectures/2009/10/23/corpus.html>
- Полевая лингвистическая практика: Учебно-методический комплекс сложной структуры. Ч. 1. Теоретические основы и методика сбора лингвистических данных для представления их в речевом корпусе русского языка / Отв. ред. А.С. Асиновский. СПб, 2007.
- Степанова С.Б., Асиновский А.С., Богданова Н.В., Русакова М.В., Шерстинова Т.Ю. 2008. Звуковой корпус русского языка повседневного общения «Один речевой день»: Концепция и состояние. *Труды международной конференции «Диалог 2008»*. М. 488–495.
- Шаров С. А. 2003. Представительный корпус русского языка в контексте мирового опыта. *Научно-техническая информация*. Сер. 2, 2003, № 6. 9–18.

Colleen R. 2004. An introduction to speech corpora. Stanford.

Leech G., Johansson S. The coming of ICAME. In: *ICAME Journal*, Vol. 33, 2009. 5-20.

Описание материально-технической базы, рекомендуемой для осуществления образовательного процесса по дисциплине (модулю)

Учебная аудитория с мультимедийным комплексом.

Описание материально-технической базы (в т.ч. программного обеспечения), рекомендуемой для адаптации электронных и печатных образовательных ресурсов для обучающихся из числа инвалидов и лиц с ОВЗ

Устанавливается образовательной организацией.

3.3 Методика проведения текущего контроля успеваемости и промежуточной аттестации и критерии оценивания

Для контроля усвоения данной дисциплины предусмотрен зачет. Контрольные мероприятия по текущему контролю знаний обучающихся проводятся в часы, отведенные для изучения дисциплины.

В течение семестра студентами выполняются практические и контрольные работы.

Порядок проведения зачета определяется ВУЗом.

3.4 Методические материалы для проведения текущего контроля успеваемости и промежуточной аттестации (контрольно-измерительные материалы, оценочные средства)

Примерные вопросы для самоконтроля:

1. Цели и задачи цифровой филологии.
2. Типы филологических задач. История использования компьютеров в филологии и лингвистике.
3. Лингвистические базы данных.
4. Типы языковых (лингвистических) ресурсов.
5. Лингвистическое обеспечение информационного поиска.
6. Системы создания электронных словарей.
7. История корпусной лингвистики в мире и в России.
8. Понятие корпуса. Модели классификации корпусов
9. Национальный корпус русского языка.
10. Другие корпуса русского языка.
11. Примеры применения корпусов текстов для лингвистических исследований
12. Инструменты искусственного интеллекта.
13. Дать определения терминов корпусной лингвистики (описание понятий):
 - Разметка корпусов
 - Репрезентативность
 - Метаданные
 - Корпус-менеджер
 - Tree bank
 - Лемматизация
 - Конкорданс
 - Параллельный корпус
14. Обучение с учителем

15. Нейронная сеть

Примерная тематика докладов, рефератов:

1. Информационно-поисковые языки.
2. Лингвистические электронные библиотеки.
3. Автоматический графематический анализ текстов.
4. Автоматическая морфологическая разметка текстов.
5. Методы снятия морфологической омонимии.
6. Способы использования корпусов в лексикографии.
7. WordNet. Русский WordNet.
8. Средства построения частотных словарей.
9. Система Multitran.
10. Использование корпусов в лингвистических исследованиях.
11. Многомерный анализ данных.
12. Экспертные системы.

Примерный перечень вопросов к зачету (экзамену) по всему курсу:

1. Искусственный интеллект в гуманитарной сфере: цели и задачи.
2. Филологические и лингвистические программы и ресурсы, в том числе работающие в режиме онлайн.
3. Поисковые системы Интернета с точки зрения филолога.
4. Лингвостатистика. Частотные словари.
5. Компьютерная лексикография. Системы создания электронных словарей.
6. Корпуса текстов и их роль в современной лингвистике.
7. Корпусная лингвистика в России. Корпусы русского языка.
8. Классификация (типология) корпусов.
9. Понятие разметки. Типы разметки.
10. Корпус как поисковая система (корпусные менеджеры).
11. Коллокации. Сервисы исследования биграмм и n грамм.
12. Лингвистические исследования, базирующиеся на корпусах.
13. Корпус как инструмент обучения языку.
14. Лингвистические базы данных и экспертные системы.
15. Подходы к классификации: обучение с учителем и обучение без учителя.
16. Самообучающиеся системы: нейронные сети.

3.5 Материально-техническое обеспечение

Минимально необходимый для реализации курса перечень материально-технического обеспечения включает лекционные аудитории (с компьютерным и видеопроекционным оборудованием для презентаций, средствами звуковоспроизведения и экраном, с выходом в Интернет). Количество индивидуальных рабочих станций должно соответствовать количеству студентов.

3.6 Информационное обеспечение

Рекомендуемая обязательная литература

Захаров В.П., Богданова С.Ю. 2020. Корпусная лингвистика. СПб.

- Захаров В.П., Рубинер В.И., Филиппов А.К., Хохлова М.В. 2013. Компьютерные технологии в лингвистике и филологии. Ч. 2 / Под общ. ред. В.П. Захарова: Учебно-методическое пособие. СПб: СПбГУ. РИО. Филологический факультет.
- Прикладное языкознание. Учебник. / Отв. ред. А.С. Герд. СПб.: СПбГУ, 1996.
- Прикладная и компьютерная лингвистика / Под ред. И. С. Николаева, О. В. Митрениной, Т. М. Ландо. Изд. 2-е. М.: ЛЕНАНД, 2017.

Рекомендуемая дополнительная литература

- Schreibman S., Siemens R., Unsworth J. 2004. A companion to Digital Humanities. Oxford
- Богданова С.Ю. 2011. Корпусная лингвистика. Иркутск
- Змитрович А. И. 1997. Интеллектуальные информационные системы. Мн.: НТООО ТерраСистемс.
- Лорьер Ж.-Л. 1991. Системы искусственного интеллекта. Пер. с фр. Москва, Мир.
- Ляшевская О.Н. 2016. Корпусные инструменты в грамматических исследованиях русского языка: учебное пособие. М.: ЯСК.
- Полевая лингвистическая практика: Учебно-методический комплекс сложной структуры. Ч. 1. Теоретические основы и методика сбора лингвистических данных для представления их в речевом корпусе русского языка / Отв. ред. А.С. Асиновский. СПб, 2007.
- Поляков В.Н. 2001. Методы искусственного интеллекта и экспертные системы. Нейросети и нейрокомпьютеры. (Учебное пособие). Казань: Изд-во КГУ.
- Террас М., Найхан Дж., Ванхутт Э., Кижнер И. 2017. Цифровые гуманитарные науки: хрестоматия Сибирский федеральный университет.

Рекомендуемый перечень иных информационных источников

1. АBBYY Россия. URL: <http://www.abbyy.ru/products/>
2. Czech National Corpus. URL: <http://ucnk.ff.cuni.cz/english/index.php>
3. Ethnologue. Languages of the World. URL: <http://www.ethnologue.com/>
4. ICAME Corpus Collection. URL: <http://icame.uib.no/corpora.html>
5. The ELRA catalogue. URL: <http://catalogue.elra.info/en-us/>
6. The LDC Corpus Catalog. URL: <https://catalog.ldc.upenn.edu/byyear>
7. Russian Corpora in Tübingen <https://www.lingexp.uni-tuebingen.de/sfb441/b1/en/korpora.html>
8. Атлас языков мира. URL: <http://wals.info/>
9. База данных «Языки мира». URL: https://iling-ran.ru/web/ru/news/201008_langworld
10. Древнерусские берестяные грамоты. URL: <http://gramoty.ru/>
11. Корпус русского литературного языка. URL: <http://narusco.ru/>
12. Корпуса и коллекции Марка Дэвиса. URL: 1) <https://www.mark-davies.info/> / 2) <https://www.english-corpora.org/>
13. British National Corpus. URL: <http://www.natcorp.ox.ac.uk/>
14. Машинный фонд русского языка. URL: <http://cfrl.ruslang.ru/index0.php>
15. Мультитран. URL: <http://multitrans.ru>
16. Национальный корпус русского языка. URL: <http://ruscorpora.ru/>
17. Новый частотный словарь русской лексики. URL: <http://dict.ruslang.ru/freq.php>
18. Программное обеспечение в области автоматической обработки текста. URL: <http://www.aot.ru/onlinedemo.html>
19. Сайт «Вавилонская башня». URL: <http://starling.rinet.ru/>
20. Сайт компании «Стэл — компьютерные системы». URL: <http://www.stel.ru/>
21. Сайт компании «Центр речевых технологий». URL: <http://www.speechpro.ru/>

22. Словари, созданные на основе Национального корпуса русского языка. URL: <http://dict.ruslang.ru/>
23. Упсальский корпус русского языка. URL: <https://snd.gu.se/en/catalogue/study/ext0071>
24. Хельсинский аннотированный корпус. URL: <http://h248.it.helsinki.fi/hanco/>

Раздел 4. Разработчики программы

Захаров Виктор Павлович, кандидат филологических наук, доцент кафедры математической лингвистики, СПбГУ.

Хохлова Мария Владимировна, кандидат филологических наук, доцент кафедры математической лингвистики, СПбГУ.

Рабочая группа ФУМО 45.00.00 по проблемам искусственного интеллекта в языкознании и литературоведении.

