

МИНОБРНАУКИ РОССИИ



Федеральное государственное бюджетное образовательное учреждение
высшего образования
«Российский государственный гуманитарный университет»
(ФГБОУ ВО «РГГУ»)

ОТДЕЛЕНИЕ ИНТЕЛЛЕКТУАЛЬНЫХ СИСТЕМ В ГУМАНИТАРНОЙ СФЕРЕ

Кафедра математики, логики и интеллектуальных систем в гуманитарной сфере

**ТЕХНОЛОГИИ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА В УПРАВЛЕНИИ
ДОКУМЕНТАМИ**

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

46.03.02 Документоведение и архивоведение с дополнительной квалификацией
в области интеллектуальных систем в гуманитарной сфере

Интеллектуальные системы в управлении документами

Уровень высшего образования: бакалавриат

Форма обучения очная

РПД адаптирована для лиц
с ограниченными возможностями
здоровья и инвалидов

Москва 2024

Технологии искусственного интеллекта в управлении документами

Рабочая программа дисциплины

Составитель:

к.т.н. Д.В. Стефановский

УТВЕРЖДЕНО

Протокол заседания кафедры МЛИИС

№ 9 от 04.04.2024

ОГЛАВЛЕНИЕ

1. Пояснительная записка	4
1.1. Цель и задачи дисциплины	4
1.2. Перечень планируемых результатов обучения по дисциплине, соотнесенных с индикаторами достижения компетенций	4
1.3. Место дисциплины в структуре образовательной программы	4
2. Структура дисциплины.....	4
3. Содержание дисциплины.....	5
4. Образовательные технологии	6
5. Оценка планируемых результатов обучения	6
5.1 Система оценивания	6
5.2 Критерии выставления оценки по дисциплине.....	7
5.3 Оценочные средства (материалы) для текущего контроля успеваемости, промежуточной аттестации обучающихся по дисциплине	8
6. Учебно-методическое и информационное обеспечение дисциплины.....	8
6.1 Список источников и литературы	8
6.2 Перечень ресурсов информационно-телекоммуникационной сети «Интернет».	8
6.3 Профессиональные базы данных и информационно-справочные системы.....	9
7. Материально-техническое обеспечение дисциплины	9
8. Обеспечение образовательного процесса для лиц с ограниченными возможностями здоровья и инвалидов	10
9. Методические материалы.....	11
9.1 Планы семинарских/ практических/ лабораторных занятий.....	11
9.2 Методические рекомендации по подготовке письменных работ	11
9.3 Иные материалы.....	12
Приложение 1. Аннотация рабочей программы дисциплины	13

1. Пояснительная записка

1.1. Цель и задачи дисциплины

Цель учебной дисциплины «Технологии искусственного интеллекта в управлении документами» состоит в том, чтобы помочь обучающимся овладеть возможностями технологий искусственного интеллекта для решения коммуникативных и прикладных задач документоведения.

Основными задачами изучения дисциплины являются:

- изучение методов искусственного интеллекта (ИИ) для компьютерной обработки текстов в электронном виде, обеспечивающие поддержку работы документоведов при обработке информации, анализе данных и интерпретации результатов;
- формирование умения применять инструментами ИИ при работе с электронными библиотеками, поисковыми системами; справочными информационными ресурсами цифровой гуманитаристики;
- овладение практическими навыками по использованию инструментов ИИ поиску, анализу, разработке, созданию информационных ресурсов;
- овладение технологиями ИИ поиска, анализа, и извлечения мнений из различных корпусов текстов .

1.2. Перечень планируемых результатов обучения по дисциплине, соотнесённых с индикаторами достижения компетенций

Компетенции	Индикаторы компетенций	Результаты обучения по дисциплине
ОПК-3. Способен применять теоретические знания для решения поставленных задач в области документоведения и архивоведения	ОПК-3.3. Применяет теоретические знания в сфере интеллектуальных систем в профессиональной деятельности	Знать: методы классификации; архитектуру и методы обучения нейронных сетей; признаки переобучения и методы борьбы с ним. Уметь: выбирать адекватные методы решения задач классификации; применять компьютерные системы для анализа данных. Владеть: навыками обучения и применения нейронных сетей для задач обработки информации; навыками работы с системой Полианалист.
ОПК-4. Способен понимать принципы работы современных информационных технологий и использовать их для решения задач профессиональной деятельности	ОПК-4.2. Применяет современные информационно-коммуникационные технологии при решении организационно-управленческих задач в	Знать: методы классификации; архитектуру и методы обучения нейронных сетей; признаки переобучения и методы борьбы с ним. Уметь: выбирать адекватные методы решения задач классификации; применять компьютерные системы для анализа данных.

	<p>профессиональной деятельности</p>	<p>Владеть: навыками обучения и применения нейронных сетей для задач обработки информации; навыками работы с системой Полианалист.</p>
	<p>ОПК-4.3. Обладает навыками использования информационно-коммуникационных технологий в сфере документационного обеспечения управления и архивного дела</p>	<p>Знать: методы классификации; архитектуру и методы обучения нейронных сетей; признаки переобучения и методы борьбы с ним. Уметь: выбирать адекватные методы решения задач классификации; применять компьютерные системы для анализа данных. Владеть: навыками обучения и применения нейронных сетей для задач обработки информации; навыками работы с системой Полианалист.</p>
<p>ОПК-7. Способен к профессиональному росту и самосовершенствованию в области гуманитарных, социальных и лингвистических наук, а также в сфере техники и технологии информатики</p>	<p>ОПК-7.2. Пользуется современными справочными и библиотечными системами и системами дистанционного образования</p>	<p>Знать: методы классификации; архитектуру и методы обучения нейронных сетей; признаки переобучения и методы борьбы с ним. Уметь: выбирать адекватные методы решения задач классификации; применять компьютерные системы для анализа данных. Владеть: навыками обучения и применения нейронных сетей для задач обработки информации; навыками работы с системой Полианалист.</p>
	<p>ОПК-7.3. Имеет практический опыт работы с поисковыми машинами, справочными и библиотечными системами и системами дистанционного образования</p>	<p>Знать: методы классификации; архитектуру и методы обучения нейронных сетей; признаки переобучения и методы борьбы с ним. Уметь: выбирать адекватные методы решения задач классификации; применять компьютерные системы для анализа данных. Владеть: навыками обучения и применения нейронных сетей для задач обработки информации; навыками работы с системой Полианалист.</p>
<p>ОПК-8. Способен понимать принципы работы современных информационных технологий и</p>	<p>ОПК-8.1. Знает содержание, структуру и принципы работы современных</p>	<p>Знать: методы классификации; архитектуру и методы обучения нейронных сетей; признаки переобучения и методы борьбы с ним. Уметь:</p>

использовать их для решения задач профессиональной деятельности	информационных технологий, применяемых для решения задач профессиональной деятельности	выбирать адекватные методы решения задач классификации; применять компьютерные системы для анализа данных. Владеть: навыками обучения и применения нейронных сетей для задач обработки информации; навыками работы с системой Полианалист.
	ОПК-8.2. Умеет обоснованно выбирать современные информационные технологии для решения задач профессиональной деятельности	Знать: методы классификации; архитектуру и методы обучения нейронных сетей; признаки переобучения и методы борьбы с ним. Уметь: выбирать адекватные методы решения задач классификации; применять компьютерные системы для анализа данных. Владеть: навыками обучения и применения нейронных сетей для задач обработки информации; навыками работы с системой Полианалист.
	ОПК-8.3. Владеет навыками применения современных информационных технологий при решении задач профессиональной деятельности	Знать: методы классификации; архитектуру и методы обучения нейронных сетей; признаки переобучения и методы борьбы с ним. Уметь: выбирать адекватные методы решения задач классификации; применять компьютерные системы для анализа данных. Владеть: навыками обучения и применения нейронных сетей для задач обработки информации; навыками работы с системой Полианалист.
ПК-1. Способен осуществлять обработку документной информации с использованием современных математических методов и алгоритмов	ПК-1.3. способен использовать математические методы для анализа, обработки информации документов, обеспечения поиска и использования информации	Знать: методы классификации; архитектуру и методы обучения нейронных сетей; признаки переобучения и методы борьбы с ним. Уметь: выбирать адекватные методы решения задач классификации; применять компьютерные системы для анализа данных. Владеть: навыками обучения и применения нейронных сетей для задач обработки информации; навыками работы с системой Полианалист.

ПК-4 Способен осуществлять проектирование и внедрение систем электронного документооборота в организации	ПК-4.1. обеспечивает доступ пользователей и ведение информационно-справочной работы в информационной системе	Знать: методы классификации; архитектуру и методы обучения нейронных сетей; признаки переобучения и методы борьбы с ним. Уметь: выбирать адекватные методы решения задач классификации; применять компьютерные системы для анализа данных. Владеть: навыками обучения и применения нейронных сетей для задач обработки информации; навыками работы с системой Полианалист.
ПК-5. Способен организовать оперативное и архивное хранение документов с использованием интегрированной среды разработки, включая средства визуального программирования, умеет использовать средства автоматизации этапов анализа и проектирования	ПК-5.1. Знает основные требования к программной документации, зафиксированные в стандартах	Знать: методы классификации; архитектуру и методы обучения нейронных сетей; признаки переобучения и методы борьбы с ним. Уметь: выбирать адекватные методы решения задач классификации; применять компьютерные системы для анализа данных. Владеть: навыками обучения и применения нейронных сетей для задач обработки информации; навыками работы с системой Полианалист.
	ПК-5.2. Имеет практический опыт разработки и тестирования прикладных программ в области оперативного и архивного хранения документов	Знать: методы классификации; архитектуру и методы обучения нейронных сетей; признаки переобучения и методы борьбы с ним. Уметь: выбирать адекватные методы решения задач классификации; применять компьютерные системы для анализа данных. Владеть: навыками обучения и применения нейронных сетей для задач обработки информации; навыками работы с системой Полианалист.
ПК-6. Способен к участию в разработке архитектур информационных и интеллектуальных систем в управлении документами	ПК-6.2. Знает способы представления архитектуры информационных и интеллектуальных систем и примеры типичных архитектур информационных и	Знать: методы классификации; архитектуру и методы обучения нейронных сетей; признаки переобучения и методы борьбы с ним. Уметь: выбирать адекватные методы решения задач классификации; применять компьютерные системы для анализа данных. Владеть:

	интеллектуальных систем в управлении документами и архивном хранении	навыками обучения и применения нейронных сетей для задач обработки информации; навыками работы с системой Полианалист.
	ПК-6.3. Умеет применять CASE-технологии для разработки и наглядного представления архитектуры информационных и интеллектуальных систем в управлении документами и архивном хранении	Знать: методы классификации; архитектуру и методы обучения нейронных сетей; признаки переобучения и методы борьбы с ним. Уметь: выбирать адекватные методы решения задач классификации; применять компьютерные системы для анализа данных. Владеть: навыками обучения и применения нейронных сетей для задач обработки информации; навыками работы с системой Полианалист.

1.3. Место дисциплины в структуре образовательной программы

Дисциплина «Технологии искусственного интеллекта в управлении документами» относится к обязательной части блока дисциплин учебного плана.

Для освоения дисциплины необходимы знания, умения и владения, сформированные в ходе изучения следующих дисциплин и прохождения практик: Информатика, Дискретная математика, Машинное обучение.

В результате освоения дисциплины формируются знания, умения и владения, необходимые для изучения следующих дисциплин и прохождения практик: Разработка информационных систем, Управление документами в информационных системах.

2. Структура дисциплины¹

Общая трудоемкость дисциплины составляет 6 з.е., 216 академических часов.

Структура дисциплины для очной формы обучения

Объем дисциплины в форме контактной работы обучающихся с педагогическими работниками и (или) лицами, привлекаемыми к реализации образовательной программы на иных условиях, при проведении учебных занятий:

Семестр	Тип учебных занятий	Количество часов
5	Лекции	24
5	Семинары	36
6	Лекции	24
6	Семинары	36
Всего:		120

¹ При реализации образовательной программы на очно-заочной и заочной формах обучения, таблица составляется для каждой формы.

Объем дисциплины (модуля) в форме самостоятельной работы обучающихся составляет 96 академических часа(ов).

3. Содержание дисциплины²

№	Наименование раздела дисциплины	Содержание
	Тема 1 Основы обучения модели ИИ	Основы обучения и линейные модели. Обучение с учителем и параметрические функции. Обучающий, тестовый и контрольный наборы. Бинарная классификация. Логлинейная бинарная классификация. Многоклассовая классификация. Представления в виде унитарного и плотного векторов. Логлинейная многоклассовая классификация. Функции потерь. Регуляризация. Градиентная оптимизация.
	Тема 2. Многослойные перцептроны	Ограничения линейных моделей: проблема XOR. Нелинейные преобразования входа. Ядерные методы. Обучаемые отображающие функции.
	Тема 3. Нейронные сети прямого распространения	Метафора, инспирированная мозгом. Математическая нотация. Репрезентативная способность. Стандартные нелинейности. Функции потерь. Регуляризация и прореживание. Слои вычисления сходства и расстояния. Слои погружения.
	Тема 4. Обучение нейронной сети.	Прямое вычисление. Вычисление на обратном проходе (производные, обратное распространение). Программное обеспечение. Рецепт реализации. Композиция сети. Выбор алгоритма оптимизации. Инициализация. Перезапуск и ансамбли. Исчезающие и взрывающиеся градиенты. Насыщение и мертвые нейроны. Тасование. Скорость обучения.
	Тема 5. Признаки для текстовых данных.	Типология проблем классификации NLP. Признаки для проблем NLP. Непосредственно наблюдаемые свойства. Производные лингвистические свойства. Базовые и комбинационные признаки. N-граммные признаки. Дистрибутивные признаки.
	Тема 6. Примеры признаков в NLP	Классификация документов: определение языка. Классификация документов: тематическая классификация. Классификация документов: установление авторства. Слово в контексте: частеречная разметка. Слово в контексте: распознавание именованных

² Раздел может быть представлен как в текстовой форме, так и в таблице

		<p>сущностей.</p> <p>Слово в контексте, лингвистические признаки: разрешение лексической многозначности предлогов.</p> <p>Отношения между словами в контексте: анализ методом разложения на дуги.</p>
	<p>Тема 7. Основы поиска в вебе. Обход и индексирование веба. Анализ ссылок</p>	<p>Кодирование категориальных признаков.</p> <p>Унитарное кодирование.</p> <p>Плотные погружения (погружения признаков) .</p> <p>Плотные векторы и унитарные представления.</p> <p>Объединение плотных векторов.</p> <p>Оконные признаки.</p> <p>Переменное число признаков: непрерывный мешок слов.</p> <p>Соотношение между унитарным и плотным векторами .</p> <p>Дистанционные и позиционные признаки.</p> <p>Дополнение, неизвестные слова и прореживание слов.</p> <p>Комбинации признаков.</p> <p>Обобществление векторов.</p> <p>Размерность.</p> <p>Словарь погружений.</p> <p>Выход сети.</p>

4. Образовательные технологии

№ п/п	Наименование раздела	Виды учебных занятий	Образовательные технологии
1	2	3	4
1	Тема 1 Основы обучения модели ИИ	Лекция 1-4 Семинар 1-6	Лекция/семинар с использованием видеоматериалов, проекционной техники. На семинарских занятия обучение проводятся с использованием программного продукта Полианалист 6.5 или выше. Выполнения задач с преподавателем, в том числе с использованием Python. Консультирование и проверка домашних заданий через социальные сети и электронную почту.
2	Тема 2. Многослойные перцептроны	Лекция 5-6 Семинар 7-9	Лекция/семинар с использованием видеоматериалов, проекционной техники. На семинарских занятия обучение проводятся с использованием программного продукта Полианалист 6.5 или выше. Выполнения задач с

			преподавателем, в том числе с использованием Python. Консультирование и проверка домашних заданий через социальные сети и электронную почту.
3	Тема 3. Нейронные сети прямого распространения	Лекция 7-8 Семинар 10-12	Лекция/семинар с использованием видеоматериалов, проекционной техники. На семинарских занятия обучение проводятся с использованием программного продукта Полианалист 6.5 или выше. Выполнения задач с преподавателем, в том числе с использованием Python. Консультирование и проверка домашних заданий через социальные сети и электронную почту.
4	Тема 4. Обучение нейронной сети.	Лекция 9-12 Семинар 13-18	Лекция/семинар с использованием видеоматериалов, проекционной техники. На семинарских занятия обучение проводятся с использованием программного продукта Полианалист 6.5 или выше. Выполнения задач с преподавателем, в том числе с использованием Python. Консультирование и проверка домашних заданий через социальные сети и электронную почту.
5	Тема 5. Признаки для текстовых данных.	Лекция 13-16 Семинар 19-24	Лекция/семинар с использованием видеоматериалов, проекционной техники. На семинарских занятия обучение проводятся с использованием программного продукта Полианалист 6.5 или выше. Выполнения задач с преподавателем, в том числе с использованием Python. Консультирование и проверка домашних заданий через социальные сети и электронную почту.
6	Тема 6. Примеры признаков в NLP	Лекция 17-20 Семинар 25-30	Лекция/семинар с использованием видеоматериалов, проекционной техники. На семинарских занятия обучение проводятся с использованием программного продукта Полианалист 6.5 или выше. Выполнения задач с

			преподавателем, в том числе с использованием Python. Консультирование и проверка домашних заданий через социальные сети и электронную почту.
7	Тема 7. Основы поиска в вебе. Обход и индексирование веба. Анализ ссылок	Лекция 21-24 Семинар 31-36	Лекция/семинар с использованием видеоматериалов, проекционной техники. На семинарских занятиях обучение проводится с использованием программного продукта Полианалист 6.5 или выше. Выполнения задач с преподавателем, в том числе с использованием Python. Консультирование и проверка домашних заданий через социальные сети и электронную почту.

В период временного приостановления посещения обучающимися помещений и территории РГГУ для организации учебного процесса с применением электронного обучения и дистанционных образовательных технологий могут быть использованы следующие образовательные технологии:

- видео-лекции;
- онлайн-лекции в режиме реального времени;
- электронные учебники, учебные пособия, научные издания в электронном виде и доступ к иным электронным образовательным ресурсам;
- системы для электронного тестирования;
- консультации с использованием телекоммуникационных средств.

5. Оценка планируемых результатов обучения

5.1 Система оценивания³

Форма контроля	Макс. количество баллов	
	За одну работу	Всего
Текущий контроль:		
- опрос	5 баллов	30 баллов
- участие в дискуссии на семинаре	5 баллов	30 баллов
Промежуточная аттестация – зачет с оценкой/экзамен		40 баллов
Итого за семестр		100 баллов

Полученный совокупный результат конвертируется в традиционную шкалу оценок и в шкалу оценок Европейской системы переноса и накопления кредитов (European Credit Transfer System; далее – ECTS) в соответствии с таблицей:

³ Система оценивания выстраивается в соответствии с учебным планом, где определены формы промежуточной аттестации (зачёт/зачёт с оценкой/экзамен), и структурой дисциплины, где определены формы текущего контроля. Указывается распределение баллов по формам текущего контроля и промежуточной аттестации, сроки отчётности.

100-балльная шкала	Традиционная шкала		Шкала ECTS
95 – 100	отлично	зачтено	A
83 – 94			B
68 – 82	хорошо		C
56 – 67	удовлетворительно		D
50 – 55			E
20 – 49	неудовлетворительно	не зачтено	FX
0 – 19			F

5.2 Критерии выставления оценки по дисциплине

Баллы/ Шкала ECTS	Оценка по дисциплине	Критерии оценки результатов обучения по дисциплине
100-83/ A,B	отлично/ зачтено	<p>Выставляется обучающемуся, если он глубоко и прочно усвоил теоретический и практический материал, может продемонстрировать это на занятиях и в ходе промежуточной аттестации.</p> <p>Обучающийся исчерпывающе и логически стройно излагает учебный материал, умеет увязывать теорию с практикой, справляется с решением задач профессиональной направленности высокого уровня сложности, правильно обосновывает принятые решения.</p> <p>Свободно ориентируется в учебной и профессиональной литературе.</p> <p>Оценка по дисциплине выставляется обучающемуся с учётом результатов текущей и промежуточной аттестации.</p> <p>Компетенции, закреплённые за дисциплиной, сформированы на уровне – «высокий».</p>
82-68/ C	хорошо/ зачтено	<p>Выставляется обучающемуся, если он знает теоретический и практический материал, грамотно и по существу излагает его на занятиях и в ходе промежуточной аттестации, не допуская существенных неточностей.</p> <p>Обучающийся правильно применяет теоретические положения при решении практических задач профессиональной направленности разного уровня сложности, владеет необходимыми для этого навыками и приёмами.</p> <p>Достаточно хорошо ориентируется в учебной и профессиональной литературе.</p> <p>Оценка по дисциплине выставляется обучающемуся с учётом результатов текущей и промежуточной аттестации.</p> <p>Компетенции, закреплённые за дисциплиной, сформированы на уровне – «хороший».</p>
67-50/ D,E	удовлетворительно/ зачтено	<p>Выставляется обучающемуся, если он знает на базовом уровне теоретический и практический материал, допускает отдельные ошибки при его изложении на занятиях и в ходе промежуточной аттестации.</p> <p>Обучающийся испытывает определённые затруднения в применении теоретических положений при решении практических задач профессиональной направленности стандартного уровня сложности, владеет необходимыми для этого базовыми навыками и приёмами.</p> <p>Демонстрирует достаточный уровень знания учебной литературы по дисциплине.</p> <p>Оценка по дисциплине выставляется обучающемуся с учётом результатов текущей и промежуточной аттестации.</p> <p>Компетенции, закреплённые за дисциплиной, сформированы на уровне – «достаточный».</p>
49-0/ F,FX	неудовлетворительно/ не зачтено	<p>Выставляется обучающемуся, если он не знает на базовом уровне теоретический и практический материал, допускает грубые ошибки при его изложении на занятиях и в ходе промежуточной аттестации.</p> <p>Обучающийся испытывает серьёзные затруднения в применении теоретических положений при решении практических задач профессиональной направленности стандартного уровня сложности, не владеет необходимыми для этого навыками и приёмами.</p> <p>Демонстрирует фрагментарные знания учебной литературы по дисциплине.</p>

Баллы/ Шкала ECTS	Оценка по дисциплине	Критерии оценки результатов обучения по дисциплине
		Оценка по дисциплине выставляются обучающемуся с учётом результатов текущей и промежуточной аттестации. Компетенции на уровне «достаточный», закреплённые за дисциплиной, не сформированы.

5.3 Оценочные средства (материалы) для текущего контроля успеваемости, промежуточной аттестации обучающихся по дисциплине

Вопросы для подготовки к экзамену:

1. Расскажите об основах обучения и линейные модели? Приведите примеры и поясните их.
2. Расскажите об обучении с учителем и параметрических функций? Приведите примеры и поясните их.
3. Что такое обучающий, тестовый и контрольный наборы? Приведите примеры и поясните их.
4. Расскажите о бинарной классификации? Приведите примеры и поясните их.
5. Расскажите о логлинейной бинарной классификации? Приведите примеры и поясните их.
6. Расскажите о многоклассовой классификации? Приведите примеры и поясните их.
7. Расскажите о представлениях в виде унитарного и плотного векторов? Приведите примеры и поясните их.
8. Что такое логлинейная многоклассовая классификация? Приведите примеры и поясните их.
9. Расскажите о функции потерь? Приведите примеры и поясните их.
10. Что такое Регуляризация? Приведите примеры и поясните их.
11. Расскажите о градиентной оптимизации? Приведите примеры и поясните их.
12. Расскажите об ограничениях линейных моделей: проблема XOR? Приведите примеры и поясните их.
13. Расскажите о нелинейных преобразованиях входа? Приведите примеры и поясните их.
14. Что такое ядерные методы? Приведите примеры и поясните их.
15. Что такое обучаемые отображающие функции? Приведите примеры и поясните их.
16. Расскажите о математической нотации? Приведите примеры и поясните их.
17. Что такое репрезентативная способность? Приведите примеры и поясните их.
18. Расскажите о стандартных нелинейностях? Приведите примеры и поясните их.
19. Что такое функция потерь? Приведите примеры и поясните их.
20. Что такое регуляризация и прореживание? Приведите примеры и поясните их.
21. Расскажите о слоях вычисления сходства и расстояния? Приведите примеры и поясните их.
22. Расскажите о слоях погружения? Приведите примеры и поясните их.
23. Что такое прямое вычисление? Приведите примеры и поясните их.
24. Что такое вычисление на обратном проходе (производные, обратное распространение)? Приведите примеры и поясните их.
25. Расскажите о типология проблем классификации NLP? Приведите примеры и поясните их.
26. Расскажите о признаках для проблем NLP? Приведите примеры и поясните их.
27. Расскажите о непосредственно наблюдаемых свойствах? Приведите примеры и поясните их.
28. Какие вы знаете производные лингвистические свойства? Приведите примеры и поясните их.

29. Какие вы знаете базовые и комбинационные признаки? Приведите примеры и поясните их.
30. Какие вы знаете N-граммные признаки? Приведите примеры и поясните их.
31. Какие вы знаете дистрибутивные признаки? Приведите примеры и поясните их.
- 32.
33. Расскажите о определение языка документа? Приведите примеры и поясните их.
34. Расскажите о тематической классификации? Приведите примеры и поясните их.
35. Расскажите о подходах установления авторства? Приведите примеры и поясните их.
36. Что такое частеречная разметка? Приведите примеры и поясните их.
37. Как осуществляется распознавание именованных сущностей? Приведите примеры и поясните их.
38. Какие вы знаете подходы разрешения лексической многозначности предлогов? Приведите примеры и поясните их.
39. Расскажите об отношениях между словами в контексте? Приведите примеры и поясните их.
40. Как осуществляется кодирование категориальных признаков? Приведите примеры и поясните их.
41. Как осуществляется Унитарное кодирование? Приведите примеры и поясните их.
42. Что такое Плотные погружения (погружения признаков)? Приведите примеры и поясните их.
43. Что такое Плотные векторы и унитарные представления? Приведите примеры и поясните их.
44. Что такое Объединение плотных векторов? Приведите примеры и поясните их.
45. Что такое Оконные признаки? Приведите примеры и поясните их.
46. Что такое непрерывный мешок слов? Приведите примеры и поясните их.
47. Соотношение между унитарным и плотным векторами? Приведите примеры и поясните их.
48. Что такое дистанционные и позиционные признаки? Приведите примеры и поясните их.
49. Для чего нужно прореживание слов? Приведите примеры и поясните их.
50. Что такое комбинации признаков? Приведите примеры и поясните их.
51. Что такое обобществление векторов? Приведите примеры и поясните их.
52. Что такое размерность? Приведите примеры и поясните их.
53. Что такое словарь погружений? Приведите примеры и поясните их.

Образец билета

1. то такое функция потерь? Приведите примеры и поясните их.
2. Что такое комбинации признаков? Приведите примеры и поясните их..

6. Учебно-методическое и информационное обеспечение дисциплины

6.1 Список источников и литературы⁴

Основная

1. Нейросетевые методы в обработке естественного языка / пер. с англ. А. А. Слинкина. – М.: ДМК Пресс, 2019. – 282 с.: ил.
2. Беляева, Л.Н. Автоматический (машинный) перевод / Л.Н. Беляева, М.И. Откупщикова // Прикладное языкознание: учебник. — СПб.: Изд-во С.-Петербург. ун-та, 1996. — С. 360–388.

⁴ Рекомендуется включать в списки издания из ЭБС и не более 15 печатных изданий.

3. Богданов, В.В., Реферирование / В.В. Богданов // Прикладное языкознание: учебник. — СПб.: Изд-во С.-Петербург. ун-та, 1996. — С. 389–398

Дополнительная

1. Загоруйко, Н.Г. Прикладные методы анализа данных и знаний / Н.Г. Загоруйко. — Новосибирск: Изд-во Института математики, 1999.

2. Каллан, Р. Основные концепции нейронных сетей: пер. с англ. / Р. Каллан. — М.: Издательский дом «Вильямс», 2001

6.2 Перечень ресурсов информационно-телекоммуникационной сети «Интернет».

Необходимо добавить то, что необходимо для изучения дисциплины

Национальная электронная библиотека (НЭБ) www.rusneb.ru
 ELibrary.ru Научная электронная библиотека www.elibrary.ru
 Электронная библиотека Grebennikon.ru www.grebennikon.ru
 Cambridge University Press
 ProQuest Dissertation & Theses Global
 SAGE Journals
 Taylor and Francis
 JSTOR

6.3 Профессиональные базы данных и информационно-справочные системы

Доступ к профессиональным базам данных: <https://liber.rsuh.ru/ru/bases>

Информационные справочные системы:

1. Консультант Плюс
2. Гарант

7. Материально-техническое обеспечение дисциплины

Для обеспечения дисциплины используется материально-техническая база образовательного учреждения: учебные аудитории, оснащённые доской, а также компьютером и проектором для демонстрации учебных материалов.

Состав программного обеспечения:

1. Windows
2. Microsoft Office

8. Обеспечение образовательного процесса для лиц с ограниченными возможностями здоровья и инвалидов

В ходе реализации дисциплины используются следующие дополнительные методы обучения, текущего контроля успеваемости и промежуточной аттестации обучающихся в зависимости от их индивидуальных особенностей:

- для слепых и слабовидящих: лекции оформляются в виде электронного документа, доступного с помощью компьютера со специализированным программным обеспечением; письменные задания выполняются на компьютере со специализированным программным обеспечением или могут быть заменены устным ответом; обеспечивается индивидуальное равномерное освещение не менее 300 люкс; для выполнения задания при необходимости

предоставляется увеличивающее устройство; возможно также использование собственных увеличивающих устройств; письменные задания оформляются увеличенным шрифтом; экзамен и зачёт проводятся в устной форме или выполняются в письменной форме на компьютере.

- для глухих и слабослышащих: лекции оформляются в виде электронного документа, либо предоставляется звукоусиливающая аппаратура индивидуального пользования; письменные задания выполняются на компьютере в письменной форме; экзамен и зачёт проводятся в письменной форме на компьютере; возможно проведение в форме тестирования.

- для лиц с нарушениями опорно-двигательного аппарата: лекции оформляются в виде электронного документа, доступного с помощью компьютера со специализированным программным обеспечением; письменные задания выполняются на компьютере со специализированным программным обеспечением; экзамен и зачёт проводятся в устной форме или выполняются в письменной форме на компьютере.

При необходимости предусматривается увеличение времени для подготовки ответа.

Процедура проведения промежуточной аттестации для обучающихся устанавливается с учётом их индивидуальных психофизических особенностей. Промежуточная аттестация может проводиться в несколько этапов.

При проведении процедуры оценивания результатов обучения предусматривается использование технических средств, необходимых в связи с индивидуальными особенностями обучающихся. Эти средства могут быть предоставлены университетом, или могут использоваться собственные технические средства.

Проведение процедуры оценивания результатов обучения допускается с использованием дистанционных образовательных технологий.

Обеспечивается доступ к информационным и библиографическим ресурсам в сети Интернет для каждого обучающегося в формах, адаптированных к ограничениям их здоровья и восприятия информации:

- для слепых и слабовидящих: в печатной форме увеличенным шрифтом, в форме электронного документа, в форме аудиофайла.

- для глухих и слабослышащих: в печатной форме, в форме электронного документа.

- для обучающихся с нарушениями опорно-двигательного аппарата: в печатной форме, в форме электронного документа, в форме аудиофайла.

Учебные аудитории для всех видов контактной и самостоятельной работы, научная библиотека и иные помещения для обучения оснащены специальным оборудованием и учебными местами с техническими средствами обучения:

- для слепых и слабовидящих: устройством для сканирования и чтения с камерой SARA CE; дисплеем Брайля PAC Mate 20; принтером Брайля EmBraille ViewPlus;

- для глухих и слабослышащих: автоматизированным рабочим местом для людей с нарушением слуха и слабослышащих; акустический усилитель и колонки;

- для обучающихся с нарушениями опорно-двигательного аппарата: передвижными, регулируемые эргономическими партами СИ-1; компьютерной техникой со специальным программным обеспечением.

9. Методические материалы

9.1 Планы семинарских/ практических/ лабораторных занятий

Тема 1. Вопросы для обсуждения:

Расскажите об основах обучения и линейные модели?

Расскажите об обучении с учителем и параметрических функций?

Что такое обучающий, тестовый и контрольный наборы?

Расскажите о бинарной классификации?

Расскажите о логлинейной бинарной классификации?
 Расскажите о многоклассовой классификации?
 Расскажите о представлениях в виде унитарного и плотного векторов?
 Что такое логлинейная многоклассовая классификация?
 Расскажите о функции потерь?
 Что такое Регуляризация?
 Расскажите о градиентной оптимизации?

Тема 2. Вопросы для обсуждения:

Расскажите об ограничениях линейных моделей: проблема XOR?
 Расскажите о нелинейных преобразованиях входа?
 Что такое ядерные методы?
 Что такое обучаемые отображающие функции?

Тема 3. Вопросы для обсуждения:

Расскажите о математической нотации?
 Что такое репрезентативная способность?
 Расскажите о стандартных нелинейностях?
 Что такое функция потерь?
 Что такое регуляризация и прореживание?
 Расскажите о слоях вычисления сходства и расстояния?
 Расскажите о слоях погружения?

Тема 4. Вопросы для обсуждения:

Что такое прямое вычисление?
 Что такое вычисление на обратном проходе (производные, обратное распространение)?
 Что такое Выбор алгоритма оптимизации?

Тема 5. Вопросы для обсуждения:

Расскажите о типологии проблем классификации NLP?
 Расскажите о признаках для проблем NLP?
 Расскажите о непосредственно наблюдаемых свойствах?
 Какие вы знаете производные лингвистические свойства?
 Какие вы знаете базовые и комбинационные признаки?
 Какие вы знаете N-граммные признаки?
 Какие вы знаете дистрибутивные признаки?

Тема 6. Вопросы для обсуждения:

Расскажите о определении языка документа?
 Расскажите о тематической классификации?
 Расскажите о подходах установления авторства?
 Что такое частеречная разметка?
 Как осуществляется распознавание именованных сущностей?
 Какие вы знаете подходы разрешения лексической многозначности предлогов?

Тема 7. Вопросы для обсуждения:

Расскажите об отношениях между словами в контексте?
 Как осуществляется кодирование категориальных признаков?
 Как осуществляется Унитарное кодирование?
 Что такое Плотные погружения (погружения признаков)?
 Что такое Плотные векторы и унитарные представления?
 Что такое Объединение плотных векторов?

Что такое Оконные признаки?

Что такое непрерывный мешок слов?

Соотношение между унитарным и плотным векторами?

Что такое дистанционные и позиционные признаки?

Для чего нужно прореживание слов?

Что такое комбинации признаков?

Что такое обобществление векторов?

Что такое размерность?

Что такое словарь погружений?

В процессе обсуждения вопросов на семинаре студенты должны собрать корпус текстов и с использованием программного продукта Полианалист и провести извлечение именованных сущностей, определить язык документа, а также провести классификацию по темам. Результат представляется потоком работ реализованным в Полианалисте.

9.2 Методические рекомендации по подготовке письменных работ⁵

9.3 Иные материалы

Любой вид занятий, создающий условия для зарождения самостоятельной мысли, познавательной и творческой активности студента связан с самостоятельной работой. В широком смысле под самостоятельной работой понимают совокупность всей самостоятельной деятельности студентов как в учебной аудитории, так и вне ее, в контакте с преподавателем и в его отсутствие. Самостоятельная работа может реализовываться: непосредственно в процессе аудиторных занятий – на лекциях, практических и семинарских занятиях, при выполнении контрольных и лабораторных работ и др.; в контакте с преподавателем вне рамок аудиторных занятий – на консультациях по учебным вопросам, в ходе творческих контактов, при ликвидации задолженностей, при выполнении индивидуальных заданий и т.д.; в библиотеке, дома, в общежитии, на кафедре и других местах при выполнении студентом учебных и творческих заданий.

Лекции

Главное в период подготовки к лекционным занятиям – научиться методам самостоятельного умственного труда, сознательно развивать свои творческие способности и овладевать навыками творческой работы. Для этого необходимо строго соблюдать дисциплину учебы и поведения. Четкое планирование своего рабочего времени и отдыха является необходимым условием для успешной самостоятельной работы. В основу его нужно положить рабочие программы изучаемых в семестре дисциплин. Каждому студенту следует составлять еженедельный и семестровый планы работы, а также план на каждый рабочий день. С вечера всегда надо распределять работу на завтрашний день. В конце каждого дня целесообразно подводить итог работы: тщательно проверить, все ли выполнено по намеченному плану, не было ли каких-либо отступлений, а если были, по какой причине это произошло. Нужно осуществлять самоконтроль, который является необходимым условием успешной учебы. Если что-то осталось невыполненным, необходимо изыскать время для завершения этой части работы, не уменьшая объема недельного плана.

⁵ В раздел включаются требования к подготовке, содержанию, и оформлению письменных работ предусмотренных учебным планом или рабочей программой (курсовая работа, эссе, реферат, доклад и т.п.). При наличии кафедральных или факультетских рекомендаций по подготовке письменных работ раздел включается в РПД по усмотрению преподавателя - составителя. Если письменная работа не предусмотрена, раздел не включается.

Самостоятельная работа на лекции Слушание и запись лекций – сложный вид вузовской аудиторной работы. Внимательное слушание и конспектирование лекций предполагает интенсивную умственную деятельность студента. Краткие записи лекций, их конспектирование помогает усвоить учебный материал. Конспект является полезным тогда, когда записано самое существенное, основное и сделано это самим студентом. Не надо стремиться записать дословно всю лекцию. Такое

«конспектирование» приносит больше вреда, чем пользы. Запись лекций рекомендуется вести по возможности собственными формулировками. Желательно запись осуществлять на одной странице, а следующую оставлять для проработки учебного материала самостоятельно в домашних условиях. Конспект лекции лучше подразделять на пункты, параграфы, соблюдая красную строку. Этому в большой степени будут способствовать пункты плана лекции, предложенные преподавателям. Принципиальные места, определения, формулы и другое следует сопровождать замечаниями «важно», «особо важно», «хорошо запомнить» и т.п. Можно делать это и с помощью разноцветных маркеров или ручек. Лучше если они будут собственными, чтобы не приходилось присить их у однокурсников и тем самым не отвлекать их во время лекции. Целесообразно разработать собственную «маркографию» (значки, символы), сокращения слов. Не лишним будет и изучение основ стенографии. Работая над конспектом лекций, всегда необходимо использовать не только учебник, но и ту литературу, которую дополнительно рекомендовал лектор. Именно такая серьезная, кропотливая работа с лекционным материалом позволит глубоко овладеть знаниями.

Семинар и проведение опроса

Каждый студент должен начать с ознакомления с планом занятия, который отражает содержание предложенной темы. Тщательное продумывание и изучение вопросов плана основывается на проработке текущего материала лекции, а затем изучения обязательной и дополнительной литературы, рекомендованную к данной теме. На основе индивидуальных предпочтений студенту необходимо самостоятельно выбрать тему доклада по проблеме семинара и по возможности подготовить по нему презентацию. Если программой дисциплины предусмотрено выполнение практического задания, то его необходимо выполнить с учетом предложенной инструкции (устно или письменно). Все новые понятия по изучаемой теме необходимо выучить наизусть и внести в глоссарий, который целесообразно вести с самого начала изучения курса. Результат такой работы должен проявиться в способности студента свободно ответить на теоретические вопросы семинара, его выступлении и участии в коллективном обсуждении вопросов изучаемой темы, правильном выполнении практических заданий и контрольных работ.

Работа с литературными источниками.

В процессе подготовки к семинарским занятиям, студентам необходимо обратить особое внимание на самостоятельное изучение рекомендованной учебно-методической (а также научной и популярной) литературы. Самостоятельная работа с учебниками, учебными пособиями, научной, справочной и популярной литературой, материалами периодических изданий и Интернета, статистическими данными является наиболее эффективным методом получения знаний, позволяет значительно активизировать процесс овладения информацией, способствует более глубокому усвоению изучаемого материала, формирует у студентов свое отношение к конкретной проблеме. Более глубокому раскрытию вопросов способствует знакомство с дополнительной литературой, рекомендованной преподавателем по каждой теме семинарского или практического занятия, что позволяет студентам проявить свою индивидуальность в рамках выступления на данных занятиях, выявить широкий спектр мнений по изучаемой проблеме.

Сессия

Каждый учебный семестр заканчивается зачетно-экзаменационной сессией. Подготовка к зачетно-экзаменационной сессии, сдача зачетов и экзаменов является также самостоятельной работой студента. Основное в подготовке к сессии – повторение всего учебного материала дисциплины, по которому необходимо сдавать зачет или экзамен. Только тот студент успевает, кто хорошо усвоил учебный материал. Если студент плохо работал в семестре, пропускал лекции, не конспектировал, не изучал рекомендованную литературу, то в процессе подготовки к сессии ему придется не повторять уже знакомое, а заново в короткий срок изучать весь учебный материал. Все это зачастую невозможно сделать из-за нехватки времени. Для такого студента подготовка к зачету или экзамену будет трудным, а иногда и непосильным делом, а конечный результат – возможное отчисление из учебного заведения.

АННОТАЦИЯ РАБОЧЕЙ ПРОГРАММЫ ДИСЦИПЛИНЫ

Цель учебной дисциплины «Технологии искусственного интеллекта в управлении документами» состоит в том, чтобы помочь обучающимся овладеть возможностями технологий искусственного интеллекта для решения коммуникативных и прикладных задач документоведения.

Основными задачами изучения дисциплины являются:

- изучение методов искусственного интеллекта (ИИ) для компьютерной обработки текстов в электронном виде, обеспечивающие поддержку работы документоведов при обработке информации, анализе данных и интерпретации результатов;
- формирование умения применять инструментами ИИ при работе с электронными библиотеками, поисковыми системами; справочными информационными ресурсами цифровой гуманитаристики;
- овладение практическими навыками по использованию инструментов ИИ поиску, анализу, разработке, созданию информационных ресурсов;
- овладение технологиями ИИ поиска, анализа, и извлечения мнений из различных корпусов текстов .

Дисциплина направлена на формирование следующих компетенций:

ОПК-3. Способен применять теоретические знания для решения поставленных задач в области документоведения и архивоведения

ОПК-4. Способен понимать принципы работы современных информационных технологий и использовать их для решения задач профессиональной деятельности

ОПК-7. Способен к профессиональному росту и самосовершенствованию в области гуманитарных, социальных и лингвистических наук, а также в сфере техники и технологии информатики

ОПК-8. Способен понимать принципы работы современных информационных технологий и использовать их для решения задач профессиональной деятельности

ПК-1. Способен осуществлять обработку документной информации с использованием современных математических методов и алгоритмов

ПК-4. Способен осуществлять проектирование и внедрение систем электронного документооборота в организации

ПК-5. Способен организовать оперативное и архивное хранение документов с использованием интегрированной среды разработки, включая средства визуального программирования, умеет использовать средства автоматизации этапов анализа и проектирования

ПК-6. Способен к участию в разработке архитектур информационных и интеллектуальных систем в управлении документами

В результате освоения дисциплины обучающийся должен:

Знать:

- методы классификации;
- архитектуру и методы обучения нейронных сетей;
- признаки переобучения и методы борьбы с ним.

Уметь:

- выбирать адекватные методы решения задач классификации;
- применять компьютерные системы для анализа данных.

Владеть:

- навыками обучения и применения нейронных сетей для задач обработки информации;

навыками работы с системой Полианалист.

По дисциплине предусмотрена промежуточная аттестация в форме зачёта с оценкой и экзамена.

Общая трудоемкость освоения дисциплины составляет 6 зачетных единиц 216 часов.