

Документ подписан простой электронной подписью
Информация о владельце:
ФИО: Ярошенко Николай Николаевич
Должность: проректор по учебно-методической деятельности
Дата подписания: 04.06.2026 11:02
Уникальный программный ключ:
25cc77c6d2a242799b1569189212ec549db4bb3f

**Федеральное государственное бюджетное образовательное учреждение
высшего образования
Московский государственный институт культуры**

**УТВЕРЖДЕНО
Председатель УМС
Библиотечно-информационного
факультета
Боронина Н. В.**

**РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ (МОДУЛЯ)
Б1.В.12 ИНТЕЛЛЕКТУАЛЬНЫЙ АНАЛИЗ ДАННЫХ**

**Направление подготовки/специальности (код, наименование): 09.03.02
Информационные системы и технологии**

**Профиль подготовки/специализация: Информационные системы и цифровые
технологии в культуре**

Квалификация (степень) выпускника: Бакалавр

Форма обучения: очная

*(РПД адаптирована для лиц
с ограниченными возможностями
здоровья и инвалидов)*

1. ЦЕЛИ И ЗАДАЧИ ОСВОЕНИЯ ДИСЦИПЛИНЫ

Цели:

Целью освоения дисциплины является: сформировать у студентов комплекс знаний и практических навыков в области интеллектуального анализа текстовых данных и дата майнинга, включая работу с большими коллекциями текстов, извлечение знаний и применение ИИ технологий.

Задачи:

- изучить методы обработки и анализа больших текстовых коллекций;
- освоить технологии дата-майнинга для извлечения знаний из текстов;
- научиться применять алгоритмы машинного обучения для классификации и кластеризации текстов;
- приобрести навыки тематического моделирования и сентимент-анализа больших корпусов;
- развить способность интерпретировать результаты анализа текстовых данных для решения прикладных задач.

2. МЕСТО ДИСЦИПЛИНЫ В СТРУКТУРЕ ОПОП ВО

Дисциплина «Интеллектуальный анализ данных» входит в состав Блока 1 «Дисциплины (модули)» и относится к части, формируемой участниками образовательных отношений, ОПОП по направлению подготовки 09.03.02 Информационные системы и технологии, профиль - Информационные системы и цифровые технологии в культуре.

Дисциплина «Интеллектуальный анализ данных» изучается в седьмом, восьмом семестре. Входные знания, умения и компетенции, необходимые для изучения данного курса, формируются в процессе изучения таких дисциплин, как «Проектирование ИС», «Математика», «Вычислительные сети и системы» и «Современные информационные технологии и программное обеспечение». В результате освоения дисциплины формируются знания, умения и навыки, необходимые для изучения следующих дисциплин и прохождения практик.

Взаимосвязь курса с другими дисциплинами ООП способствует планомерному формированию необходимых компетенций и углубленной подготовке студентов к решению специальных практических профессиональных задач.

3. КОМПЕТЕНЦИИ ОБУЧАЮЩЕГОСЯ, ФОРМИРУЕМЫЕ В РЕЗУЛЬТАТЕ ОСВОЕНИЯ ДИСЦИПЛИНЫ

Процесс освоения дисциплины направлен на формирование компетенций (элементов следующих компетенций) в соответствии с ФГОС ВО и ОПОП ВО по данному направлению подготовки (специальности) 09.03.02 Информационные системы и технологии:

Перечень планируемых результатов обучения по дисциплине (модулю).

Компетенция (код и наименование)	Индикаторы компетенций	Результаты обучения

<p>ПК-4 Готовность к информационно-аналитической деятельности и решению задач её автоматизации, интеллектуальному анализу данных</p>	<p>ПК 4.3. Работает с большими данными в области управления культурой</p>	<p>Знает:</p> <ul style="list-style-type: none"> • основные методы обработки естественного языка (NLP) для больших данных; • технологии дата майнинга текстовых коллекций; • алгоритмы тематического моделирования (LDA, NMF) и кластеризации текстов; • инструменты для работы с большими текстовыми корпусами (Spark NLP, Gensim); • методы sentiment анализа и извлечения именованных сущностей (NER); • принципы работы нейросетевых моделей для анализа текстов (BERT, GPT).
		<p>Умеет:</p> <ul style="list-style-type: none"> • проводить предобработку больших текстовых коллекций (токенизация, лемматизация, удаление стоп слов); • применять методы дата майнинга для выявления закономерностей в текстах; • строить тематические модели и интерпретировать их результаты; • выполнять sentiment анализ отзывов и комментариев; • визуализировать результаты анализа текстовых данных; • использовать ИИ инструменты для автоматизации обработки текстов.
		<p>Владеет:</p> <ul style="list-style-type: none"> • навыками работы с библиотеками NLP (NLTK, SpaCy, Gensim, Transformers); • методами тематического моделирования и кластеризации текстов; • инструментами дата майнинга (Scikit learn, Spark NLP); • приемами работы с большими текстовыми корпусами и API ИИ сервисов.

4. СТРУКТУРА И СОДЕРЖАНИЕ ДИСЦИПЛИНЫ (модуля)

4.1 Объем дисциплины (модуля)

Объем (общая трудоемкость) дисциплины «Основы естественно-научных и инженерных знаний» составляет 4 з.е., 144 акад. часов, из них:

Контактных: 80 акад.ч.

СРС: 64 acad.ч.

Форма контроля: зачет, зачет с оценкой.

4.2. Структура дисциплины для очной формы обучения.

	Раздел дисциплины	Семестр	Виды учебной работы, включая самостоятельную работу студентов и трудоемкость (в часах) в т.ч. в интерактивной форме					Формы текущего контроля успеваемости (по неделям семестра) Форма промежуточной аттестации (по семестрам)
			Лекции	Семинары/ практические	Консультации	ИКР	СРС	
1	Введение в анализ текстовых данных и дата майнинг.	7	2	5		2	8	Устные опросы на лекциях Проверка отчётов по практическим работам
2	Предобработка больших текстовых коллекций.	7	2	5		2	8	Мини-тесты по темам Проверка отчётов по практическим работам
3	Методы дата майнинга для текстов.	7	2	5		2	8	Устные опросы на лекциях Проверка отчётов по практическим работам
4	Тематическое моделирование больших корпусов.	7	2	7		4	8	Защита индивидуальных проектов (тематическая модель или дашборд с аналитикой текстов)
5	Сентимент анализ и извлечение информации.	8	2	5		2	8	Устные опросы на лекциях Проверка отчётов по практическим работам
6	Нейросетевые методы для	8	2	5		2	8	Мини-тесты по темам Проверка отчётов по практическим работам

	анализа текстов.							
7	Дата майнинг в социальных медиа.	8	2	5		2	8	Мини-тесты по темам Проверка отчётов по практическим работам
8	Комплексный анализ текстовых данных.	8	2	7		4	8	Защита индивидуальных проектов (тематическая модель или дашборд с аналитикой текстов)
	Форма итогового контроля							Экзамен Зачет с оценкой
	Всего 144 час		16	44		20	64	

4.3. Содержание разделов дисциплины (модуля)

№	Наименование (подраздела, дисциплины)	раздела темы)	Содержание
1	Введение в анализ текстовых данных и дата майнинг.		<ul style="list-style-type: none"> • Понятие дата майнинга и его применение к текстовым данным. • Особенности работы с большими текстовыми коллекциями. • Источники текстовых данных: веб-страницы, соцсети, архивы, оцифрованные книги. • Примеры использования дата майнинга в сфере культуры (анализ отзывов, изучение коллекций).
2	Предобработка больших текстовых коллекций.		<ul style="list-style-type: none"> • Токенизация, лемматизация, стемминг для больших данных. • Удаление стоп слов, нормализация текста. • Векторное представление текстов: TF-IDF, Word2Vec, Doc2Vec. • Практическое занятие: предобработка корпуса отзывов о музеях (NLTK/SpaCy).
3	Методы дата майнинга для текстов.		<ul style="list-style-type: none"> • Извлечение ассоциативных правил из текстов. • Поиск частых последовательностей и шаблонов. • Кластеризация документов: K-means, иерархическая кластеризация. • Практическое занятие: кластеризация статей о культурных событиях (Gensim).

4	Тематическое моделирование больших корпусов.	<ul style="list-style-type: none"> Алгоритмы LDA и NMF для выявления тем. Оценка качества тематических моделей. Визуализация тем (pyLDAvis). Практическое занятие: построение тематической модели коллекции статей о культуре (Gensim + LDA).
5	Сентимент анализ и извлечение информации.	<ul style="list-style-type: none"> Методы определения тональности текстов. Извлечение именованных сущностей (NER): люди, места, организации. Распознавание событий и фактов в текстах. Практическое занятие: анализ тональности отзывов о выставках (SpaCy + Transformers).
6	Нейросетевые методы для анализа текстов.	<ul style="list-style-type: none"> Трансформеры (BERT, RoBERTa) для классификации и кластеризации. Fine tuning предобученных моделей. Практическое занятие: классификация статей о культуре с помощью BERT (Hugging Face).
7	Дата майнинг в социальных медиа.	<ul style="list-style-type: none"> Сбор данных из соцсетей (Twitter, VK) через API. Анализ хештегов и трендов. Выявление влиятельных пользователей и сообществ. Практическое занятие: дата майнинг твитов о культурных событиях (Tweepy + NLTK).
8	Комплексный анализ текстовых данных.	<ul style="list-style-type: none"> Интеграция методов дата майнинга и машинного обучения. Построение дашбордов с аналитикой текстов. Практическое занятие: разработка дашборда в Power BI (анализ отзывов + темы + тональность).

5. ОБРАЗОВАТЕЛЬНЫЕ ТЕХНОЛОГИИ

№ п/п	Наименование раздела	Виды учебных занятий	Образовательные технологии
1	2	3	4
I.	Введение в анализ текстовых данных и дата майнинг.	Занятие лекционного типа и практическое занятие	Вводная лекция с мультимедийным сопровождением.

			Обсуждение результатов практических работ. Самостоятельная работа студентов.
2.	Предобработка больших текстовых коллекций.	Занятие лекционного типа и практическое занятие	Обзорная лекция с мультимедийным сопровождением. Обсуждение результатов практических работ. Самостоятельная работа студентов.
3.	Методы дата майнинга для текстов.	Занятие лекционного типа и практическое занятие	Проблемная лекция с мультимедийным сопровождением. Обсуждение результатов практических работ. Самостоятельная работа студентов.
4.	Тематическое моделирование больших корпусов.	Занятие лекционного типа и практическое занятие	Лекция-визуализация. Обсуждение результатов практических работ. Защита индивидуальных проектов. Самостоятельная работа студентов.
5.	Сентимент анализ и извлечение информации.	Занятие лекционного типа и практическое занятие	Лекция-беседа. Обсуждение результатов практических работ. Самостоятельная работа студентов.
6.	Нейросетевые методы для анализа текстов.	Занятие лекционного типа	Лекция с элементами самостоятельной работы.

		и практическое занятие	Обсуждение результатов практических работ.
7.	Дата майнинг в социальных медиа.	Занятие лекционного типа и практическое занятие	Лекция с элементами самостоятельной работы. Обсуждение результатов практических работ.
8.	Комплексный анализ текстовых данных.	Занятие лекционного типа и практическое занятие	Лекция с элементами обсуждения. Обсуждение результатов практических работ. Защита индивидуальных проектов.

6. ОЦЕНОЧНЫЕ СРЕДСТВА ДЛЯ ТЕКУЩЕГО КОНТРОЛЯ УСПЕВАЕМОСТИ, ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ ПО ИТОГАМ ОСВОЕНИЯ ДИСЦИПЛИНЫ

Текущий контроль выполнения заданий (контроль формирования компетенций) осуществляется регулярно, начиная с первой недели семестра (входящий контроль). Текущий контроль освоения отдельных разделов дисциплины осуществляется при помощи опроса, заданий (проектных и практических) и тестового материала в завершении изучения каждого раздела. Система текущего контроля успеваемости служит не только оценке уровня компетентностной подготовки обучающегося и способствует в дальнейшем наиболее качественному и объективному оцениванию его в ходе промежуточной аттестации, но и самооценке обучающегося, стимулируя его усилия.

Промежуточная аттестация проводится в форме зачета, зачета с оценкой:

- теоретический вопрос (оценка знаний методов дата майнинга и NLP);
- практическое задание (анализ большого текстового корпуса);
- защита проекта (тематическая модель коллекции текстов или дашборд с аналитикой отзывов).

6.1. Система оценивания

Форма контроля	Компетенция	Оценка
Текущий контроль: - опрос - выполнение практических работ	ПК-4.3.	зачтено/не зачтено зачтено/не зачтено

Промежуточная аттестация Экзамен Зачет оценкой	ПК-4.3.	отлично/хорошо/удовлетворительно/неудовлетворительно зачтено (отлично, хорошо, удовлетворительно)/ не зачтено
---	---------	--

6.2. Критерии оценки результатов по дисциплине

Оценка по дисциплине	Критерии оценки результатов обучения по дисциплине
«отлично»/ «зачтено (отлично)»	<p>Выставляется обучающемуся, если компетенция(ии), закрепленная за дисциплиной, сформирована (по индикаторам/ результатам обучения в формате знать-уметь-владеть) в полном объеме на уровне «высокий», и обучающийся демонстрирует как результат обучения следующие знания, умения и навыки: обучающийся глубоко и прочно усвоил теоретический и практический материал, продемонстрировал это на занятиях и в ходе промежуточной аттестации.</p> <p>Обучающийся исчерпывающе и логически стройно излагает учебный материал, умеет сочетать 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>
«хорошо»/ «зачтено (хорошо)»	<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>
«удовлетворительно»/ «зачтено (удовлетворительно)»	<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>

Оценка по дисциплине	Критерии оценки результатов обучения по дисциплине
	<p>Демонстрирует достаточный уровень знания учебной литературы по дисциплине.</p> <p>Оценка по дисциплине выставляются обучающемуся с учётом результатов текущей и промежуточной аттестации.</p> <p>Компетенции, закреплённые за дисциплиной, сформированы на уровне «достаточный».</p>
<p>«неудовлетворительно»/ «не зачтено» (неудовлетворительно)</p>	<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> <p>Оценка по дисциплине выставляются обучающемуся с учётом результатов текущей и промежуточной аттестации.</p> <p>Компетенции на уровне «достаточный», закреплённые за дисциплиной, не сформированы.</p>

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

Тест по дисциплине:

Вариант 1

Часть 1. Закрытые вопросы (выберите один правильный ответ)

1. Какой метод используется для векторного представления текстов на основе частоты терминов?
 - а) PCA;
 - б) TF-IDF;
 - в) K-means;
 - г) LDA.
2. Какой инструмент подходит для лемматизации русских текстов?
 - а) OpenCV;
 - б) SpaCy;
 - в) Matplotlib;
 - г) TensorFlow.
3. Что такое сентимент-анализ?
 - а) определение тональности текста;
 - б) извлечение ключевых слов;
 - в) классификация документов;
 - г) построение тематических моделей.
4. Какой алгоритм применяется для тематического моделирования текстовых коллекций?

- а) SVM;
 - б) LDA;
 - в) CNN;
 - г) KNN.
5. Какой инструмент используется для визуализации результатов тематического моделирования?
- а) pyLDAvis;
 - б) BeautifulSoup;
 - в) NumPy;
 - г) Requests.

Часть 2. Открытые вопросы

- 6. Перечислите 3 этапа предобработки текстовых данных перед анализом.
- 7. Опишите кратко принцип работы алгоритма LDA (не более 50 слов).
- 8. Приведите пример практического применения сентимент-анализа в культурной сфере (1–2 предложения).
- 9. Назовите 2 инструмента для сбора данных из социальных сетей через API.
- 10. Какие метрики используются для оценки качества тематической модели? Назовите 2 метрики.

Ключи к варианту 1

Закрытые вопросы:

1 — б; 2 — б; 3 — а; 4 — б; 5 — а.

Открытые вопросы:

- 6. Токенизация, лемматизация/стемминг, удаление стоп-слов.
- 7. LDA (Latent Dirichlet Allocation) — вероятностная модель, которая представляет документы как смеси тем, а темы — как распределения по словам.
- 8. Анализ отзывов о выставках для определения общего настроения посетителей и выявления проблемных моментов.
- 9. Tweepy (для Twitter), VK API (для ВКонтакте).
- 10. Coherence score, perplexity.

Вариант 2

Часть 1. Закрытые вопросы (выберите один правильный ответ)

- 1. Какой метод преобразует слова в векторные представления с учётом семантики?
 - а) One-Hot Encoding;
 - б) Word2Vec;
 - в) TF-IDF;
 - г) PCA.
- 2. Что такое NER в контексте обработки текстов?
 - а) метод кластеризации;
 - б) извлечение именованных сущностей;
 - в) алгоритм классификации;
 - г) способ визуализации.
- 3. Какой инструмент используется для работы с трансформерами (BERT, GPT)?
 - а) NLTK;
 - б) Hugging Face;

- в) Gensim;
 - г) OpenCV.
4. Что означает метрика coherence score?
- а) точность классификации;
 - б) качество тематической модели;
 - в) скорость обработки данных;
 - г) размер корпуса текстов.
5. Какой метод подходит для кластеризации коротких текстов (например, твитов)?
- а) ARIMA;
 - б) K-means;
 - в) SVM;
 - г) ANOVA.

Часть 2. Открытые вопросы

- 6. Назовите 3 проблемы при работе с большими текстовыми коллекциями.
- 7. Кратко опишите принцип работы Word2Vec (не более 50 слов).
- 8. Приведите пример использования тематического моделирования в анализе литературных произведений (1–2 предложения).
- 9. Какие инструменты можно использовать для построения дашбордов с аналитикой текстов? Назовите 2 инструмента.
- 10. Что такое fine-tuning в контексте нейросетевых моделей? Кратко объясните (30–50 слов).

Ключи к варианту 2

Закрытые вопросы:

1 — б; 2 — б; 3 — б; 4 — б; 5 — б.

Открытые вопросы:

- 6. Дублирование данных, шум в текстах, высокая размерность векторных представлений.
- 7. Word2Vec — метод обучения векторных представлений слов на основе их контекста: слова с похожим значением имеют близкие векторы.
- 8. Выявление ключевых тем в романах Достоевского для анализа эволюции его творчества.
- 9. Power BI, Tableau.
- 10. Fine-tuning — дообучение предобученной модели на специфическом датасете для повышения качества решения конкретной задачи.

Примерный перечень вопросов к экзамену:

Конечно! Вот расширенный список из 20 экзаменационных вопросов — исходные 10 вопросов разбиты на более детальные, с добавлением уточняющих и смежных тем:

- 1. Что такое интеллектуальный анализ данных (ИАД)? Охарактеризуйте его ключевые задачи и отличия от традиционного анализа данных.
- 2. В чём специфика применения ИАД к текстовым данным? Приведите 3–4 примера реальных бизнес-задач, решаемых с помощью ИАД текстовых данных.
- 3. Перечислите и кратко опишите основные этапы предобработки текстовых данных перед анализом.
- 4. Что такое токенизация? Опишите подходы к токенизации для языков с пробелами (английский) и без пробелов (китайский).

5. В чём разница между стеммингом и лемматизацией? Приведите примеры для русского и английского языков. Назовите популярные стеммеры и лемматизаторы.
6. Объясните принцип работы метода TF-IDF. Как рассчитывается IDF? Приведите формулу и пример расчёта для небольшого текста.
7. Что такое Word2Vec? Опишите архитектуру Skip-gram и CBOW, укажите их сильные и слабые стороны.
8. В чём отличие Doc2Vec от Word2Vec? Какие существуют варианты реализации Doc2Vec (PV-DM, PV-DBOW)?
9. Сравните TF-IDF, Word2Vec и Doc2Vec по следующим критериям: интерпретируемость, размерность вектора, учёт контекста, вычислительная сложность. Представьте результат в виде таблицы.
10. Опишите алгоритм K-means для кластеризации текстов. Как выбрать оптимальное число кластеров (метод локтя, silhouette score)?
11. Что такое иерархическая кластеризация? Постройте дендрограмму для небольшого набора текстов и объясните, как по ней определить кластеры.
12. Что такое тематическое моделирование? В чём суть вероятностной модели LDA? Какие гиперпараметры есть у LDA (α , β) и как они влияют на результат?
13. Как интерпретировать результаты LDA? Приведите пример темы (топ-слов) и объясните, как понять, что тема осмысленная.
14. Назовите метрики для оценки качества тематических моделей (perplexity, coherence score). В чём их преимущества и недостатки?
15. Как визуально интерпретировать результаты тематического моделирования? Опишите, какие паттерны в pyLDAvis говорят о хорошей/плохой модели.
16. Перечислите основные источники текстовых данных для анализа. Укажите плюсы и минусы каждого (веб-страницы, соцсети, архивы, оцифрованные книги).
17. Какие юридические и этические ограничения нужно учитывать при сборе текстовых данных из соцсетей и веб-страниц?
18. Какие проблемы возникают при работе с большими текстовыми коллекциями (миллионы документов)? Опишите 2–3 способа их решения (распараллеливание, выборка, инкрементальное обучение).
19. Сравните NLTK и SpaCy для предобработки текстов: скорость, функциональность, поддержка языков, удобство API. Приведите пример кода токенизации на обоих инструментах.
20. Опишите функционал pyLDAvis. Как с его помощью оценить качество тематической модели и объяснить результаты не-эксперту?

Примерный перечень вопросов для к зачету с оценкой:

1. Сентимент-анализ: методы определения тональности текстов, метрики оценки качества.
2. Извлечение именованных сущностей (NER): задачи, алгоритмы, применение в культурной сфере.
3. Нейросетевые методы для анализа текстов: трансформеры (BERT, RoBERTa), их преимущества перед традиционными методами.
4. Fine-tuning предобученных языковых моделей: назначение и процесс.
5. Анализ социальных медиа: сбор данных через API, анализ хештегов и трендов.
6. Дата-майнинг в социальных сетях: выявление влиятельных пользователей и сообществ.
7. Комплексный анализ текстовых данных: интеграция методов дата-майнинга и машинного обучения.
8. Построение дашбордов с аналитикой текстов: ключевые показатели и визуализация.

9. Этические и правовые аспекты анализа текстовых данных (ФЗ-152, ФЗ-149).
10. Применение ИИ-инструментов (Yandex GPT и др.) для автоматизации анализа текстов: возможности и ограничения.

Примеры практико-ориентированных задач к зачёту с оценкой

Задача 1. Сентимент-анализ

Дан набор из 5 коротких отзывов о библиотеке.

1. Проведите ручную разметку тональности каждого отзыва: *положительный*, *отрицательный*, *нейтральный*.
2. Для одного из отзывов приведите 2–3 лексических маркера (слова/фразы), по которым вы определили тональность.
3. Назовите две метрики качества для оценки модели сентимент-анализа и кратко поясните, что они измеряют.

Пример отзыва: «Библиотекари скучные, но обстановка красивая».

Задача 2. Извлечение именованных сущностей (NER)

Дан текст на русском языке о культурном событии (фестиваль, выставка и т. д.).

1. Выделите и классифицируйте именованные сущности по типам: *PER* (персона), *ORG* (организация), *LOC* (место), *DATE* (дата).
2. Кратко опишите, как результаты NER могут помочь в анализе культурной сферы (приведите 1–2 примера применения).

Пример текста: «В Третьяковской галерее 15 мая открылась выставка работ Ильи Репина».

Задача 3. Нейросетевые методы для анализа текстов

Сравните традиционный метод TF-IDF + классификатор (например, SVM) с моделью BERT для задачи классификации новостей по темам.

1. В таблице из 3 строк укажите:
 - **Учёт контекста** (да/нет, кратко поясните);
 - **Обработка омонимии** (как решается);
 - **Требования к объёму данных** (большой/малый).

Задача 4. Fine-tuning предобученных моделей

Опишите пошаговый процесс fine-tuning модели RuBERT для классификации отзывов о музеях (классы: *положительный*, *отрицательный*).

1. Перечислите 4 ключевых шага (от подготовки данных до вывода).
2. Укажите, какие слои модели обычно замораживают/размораживают на этапе fine-tuning и почему.

Задача 5. Анализ социальных медиа

Вы собираете данные о хештегах #МоскваБиблиотека через API VK за последнюю неделю.

1. Составьте список из 5 метрик для анализа трендов по этому хештегу (например, *количество постов в день*).
2. Предложите способ визуализации этих метрик (тип графика + ось X/Y).

Задача 6. Дата-майнинг в соцсетях

Дано: граф взаимодействий пользователей в соцсети (вершины — пользователи, рёбра — подписки/лайки).

1. Назовите два алгоритма для выявления влиятельных пользователей (лидеров мнений) и кратко опишите принцип их работы (1–2 предложения на алгоритм).
2. Приведите пример метрики влиятельности (например, PageRank) и объясните, что она измеряет.

Задача 7. Комплексный анализ текстовых данных

Опишите последовательность этапов для анализа отзывов о мобильном приложении, объединяющий методы дата-майнинга и машинного обучения.

1. Перечислите 3 этапа обработки (от сырых текстов до выводов).
2. Для каждого этапа укажите:
 - метод дата-майнинга/ML;
 - цель этапа (что он даёт для общего анализа).

Задача 8. Построение дашбордов с аналитикой текстов

Вам нужно создать дашборд для мониторинга тональности упоминаний бренда в соцсетях.

1. Выберите 4 ключевых показателя (KPIs) для отображения (например, *доля положительных упоминаний*).
2. Для каждого KPI предложите тип визуализации (столбчатая диаграмма, линейный график, круговая диаграмма и т. д.) и кратко обоснуйте выбор.

Задача 9. Этические и правовые аспекты

Вы выявляет случаи нелегитимного использования личных сообщений пользователей в мессенджере для исследования общественного мнения.

1. Укажите два положения ФЗ 152-ФЗ («О персональных данных») и два положения ФЗ 149-ФЗ («Об информации...»), которые необходимо учесть.
2. Укажите два способа анонимизации данных, которые используют недобросовестные исследователи перед анализом, чтобы снизить риски нарушения закона.

Задача 10. Применение ИИ-инструментов

Используйте Yandex GPT для генерации краткого резюме (3–4 предложения) по тексту новости (предоставлен экзаменатором).

1. Скопируйте запрос к модели и получившееся резюме.
2. Оцените качество результата по двум критериям: *точность фактов, полнота*. Приведите один пример ошибки/упущения, если они есть.
3. Кратко (1–2 предложения) укажите, в каких задачах анализа текстов такие инструменты наиболее полезны, а в каких — ненадёжны.

7. УЧЕБНО-МЕТОДИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

7.1. Список литературы и источников

Основная:

1. Bird S., Klein E., Loper E. «Natural Language Processing with Python».
2. Jurafsky D., Martin J.H. «Speech and Language Processing».
3. Raschka S., Mirjalili V. «Python Machine Learning».
4. Manning C.D., Raghavan P., Schütze H. «Introduction to Information Retrieval».

Дополнительная:

1. Grimmer J., Stewart B.M. «Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts».
2. Паклин Н., Орешков В. «Бизнес аналитика: от данных к знаниям».
3. Silge J., Robinson D. «Text Mining with R: A Tidy Approach».

7.2. Перечень ресурсов информационно-телекоммуникационной сети «Интернет».

Культура РФ: <https://www.culture.ru/> [Электронный ресурс]: сайт (дата обращения 19.12.25).
Официальная документация NLTK, SpaCy, Gensim, Hugging Face.
Открытые датасеты текстов (Kaggle, Hugging Face, Common Crawl).
Датасеты отзывов (Yelp, TripAdvisor, музейные форумы).
API Twitter, VK для сбора данных.

Доступ в ЭБС:

ООО «Электронное издательство ЮРАЙТ».

8. МЕТОДИЧЕСКИЕ УКАЗАНИЯ ПО ОСВОЕНИЮ ДИСЦИПЛИНЫ (МОДУЛЯ)

8.1. Планы семинарских/практических занятий

Практические задания:

Задание 1. Литературный анализ

- Задача: провести тематическое моделирование сборника стихов русских поэтов Серебряного века.
- Данные: корпус текстов стихотворений (Блок, Ахматова, Маяковский и др.).
- Инструменты: Python, Gensim (LDA), NLTK/SpaCy.
- Результат: визуализация тем и ключевых слов для каждой темы, сравнение тематических предпочтений поэтов.

Задание 2. Политический анализ

- Задача: проанализировать тональность твитов о культурной политике России за последний год.
- Данные: твиты с хештегом #культурнаяполитика (через Twitter API).
- Инструменты: Python, TextBlob/VADER (сентимент-анализ), Matplotlib (визуализация).
- Результат: график изменения тональности обсуждений во времени, облако тегов наиболее обсуждаемых тем.

Задание 3. Анализ культурной политики

- Задача: сравнить ключевые темы в государственных программах поддержки культуры 2010 и 2020 годов.
- Данные: тексты программ с официальных сайтов Минкультуры.
- Инструменты: Python, TF-IDF, Word2Vec.
- Результат: сравнительная таблица ключевых тем, визуализация различий в приоритетах.

Задание 4. Анализ читательских предпочтений

- Задача: выявить популярные жанры в региональной библиотеке за последние 5 лет.
- Данные: статистика выдачи книг (жанр, автор, год издания).
- Инструменты: Power BI, Pandas (агрегация данных).
- Результат: дашборд с динамикой популярности жанров, топ-10 авторов по годам.

Задание 5. Мониторинг культурного разнообразия

- Задача: проанализировать репертуар театров Москвы на представленность современных российских авторов.
- Данные: афиши театров за сезон (парсинг с сайтов).
- Инструменты: Python (BeautifulSoup), SpaCy (NER для извлечения имён авторов).
- Результат: процент пьес современных российских авторов в репертуаре, топ-5.

Практические задания (на выбор):

1. Провести предобработку корпуса отзывов о культурных мероприятиях (удалить стоп-слова, привести к нижнему регистру, выполнить лемматизацию).
2. Построить TF-IDF-матрицу для небольшого корпуса текстов (5–10 документов).
3. Выполнить кластеризацию коротких текстов (например, твитов о музеях) с помощью алгоритма K-means.
4. Построить простую тематическую модель (LDA) для небольшого набора статей о культуре (3–5 тем).
5. Визуализировать результаты тематического моделирования с помощью облака слов для каждой темы.

6. Извлечь ключевые слова из текста с помощью TF-IDF.
7. Сравнить два текста на схожесть с использованием векторных представлений (косинусное сходство).
8. Подготовить отчёт с описанием этапов предобработки и анализа текстового корпуса (500–700 слов).
9. Выполнить sentiment-анализ корпуса отзывов о выставках (определить общую тональность, построить график распределения положительных/отрицательных отзывов).
10. Извлечь именованные сущности (люди, места, организации) из статей о культурных событиях с помощью SpaCy.
11. Провести тематическое моделирование большого корпуса статей о культурной политике (100+ документов, 5–7 тем) с использованием LDA или NMF.
12. Классифицировать тексты (например, новости о театре, музыке, изобразительном искусстве) с помощью модели BERT (Hugging Face).
13. Собрать данные из соцсетей (например, твиты с хештегом #культурнаяполитика) через API и выполнить анализ трендов (частота хештегов, ключевые слова).
14. Построить дашборд в Power BI или Tableau с визуализацией результатов анализа текстов (тональность, темы, ключевые слова, динамика обсуждений).
15. Выполнить fine-tuning модели BERT на корпусе текстов о культуре для задачи классификации (например, «актуальные события» vs «архивные материалы»).
16. Разработать сценарий использования ИИ-ассистента (на базе Yandex GPT) для анализа запросов посетителей культурного учреждения и генерации отчётов.
17. Оценить качество модели тематического моделирования (LDA) с использованием метрик coherence score и perplexity.
18. Подготовить итоговый отчёт (800–1000 слов) с описанием комплексного анализа текстового корпуса: этапы, методы, результаты, выводы и рекомендации.

Рекомендации по работе с литературой

- 1) Использовать рекомендованные источники, а также все возможные и доступные источники, содержащие материал.
- 2) Конспектирование первоисточников предполагает краткое, лаконичное письменное изложение основного содержания, смысла (доминанты) какого-либо текста. Вместе с тем этот процесс требует активной мыслительной работы. Конспектируемый материал содержит информацию трех видов: главную, второстепенную и вспомогательную. Главной является информация, имеющая основное значение для раскрытия сущности того или иного вопроса, темы. Второстепенная информация служит для пояснения, уточнения главной мысли, например, описание деталей быта, одежды, характеризующих культуру народа. К этому типу информации относятся разного рода комментарии (объяснительные замечания, толкования). Назначение вспомогательной информации – помочь читателю лучше понять данный материал. Это всякого рода напоминания о ранее излагавшемся материале, заголовки, вопросы.
- 3) Работая над текстом автора, следует избегать механического переписывания текста. Важно выделять главные положения, фиксирование которых сопровождается, в случае необходимости, цитатами. Вспомогательную информацию при конспектировании не записывают. Усвоению нового материала неоценимую помощь оказывают собственные схемы, рисунки, таблицы, графическое выделение важной мысли. На каждой странице конспекта возможно выделение трех-четырех важных моментов по определенной теме. Необходимо в конспекте отражать сущность проблемы, поставленного вопроса, что служит решению поставленной на семинаре задачи.

8.2. Методические рекомендации к самостоятельной работе студентов

Примерные направления для самостоятельной работы:

- изучение документации NLTK, SpaCy и Gensim;
- сбор и предобработка текстового корпуса (отзывы, статьи о культуре);
- работа с API OpenAI/Yandex GPT для генерации отчетов;
- обучение нейросетевой модели для классификации текстов.

Самостоятельная работа обучающихся включает в себя такие виды и формы как: подготовка к практическому занятию, подготовка к дискуссии, презентации, подготовка доклада, конспектирование изучаемой литературы, аналитический обзор новой литературы по изучаемой теме, написание эссе и др.

Для более углубленного изучения материала задание для самостоятельной работы рекомендуется выполнять параллельно с изучением данной темы. При выполнении заданий для самостоятельной работы, по возможности, следует ориентироваться на наглядное представление материала.

8.4. Иные материалы

Подготовка к дискуссии

Дискуссия - это обсуждение какого-либо спорного вопроса, проблемы. Важной характеристикой дискуссии, отличающей ее от других видов спора, является аргументированность.

Дискуссия предполагает выработку и активное продвижение своей точки зрения по изучаемой проблеме, умение выслушать альтернативную точку зрения, вступить в полемику, на основе изложения и учета всех точек зрения прийти к объективному результату.

Основные шаги при подготовке к дискуссии.

Выбор темы дискуссии определяется целями обучения и содержанием учебного материала. При этом на обсуждение обучающихся выносятся темы, имеющие проблемный характер, содержащие в себе противоречивые точки зрения, дилеммы, задевающие привычные установки обучающихся. Тема разбивается на отдельные вопросы, которые сообщаются обучающимся. Указывается литература, справочные материалы, необходимые для подготовки к дискуссии.

Проведение дискуссии:

формулирование проблемы и целей дискуссии;

создание мотивации к обсуждению - определение значимости проблемы, указание на нерешенность и противоречивость вопроса и т.д.;

установление регламента дискуссии и ее основных этапов;

совместная выработка правил дискуссии;

выяснение однозначности понимания темы дискуссии, используемых в ней терминов, понятий.

Приемы введения в дискуссию:

предъявление проблемной ситуации;

демонстрация видеосюжета;

демонстрация материалов (статей, документов);

ролевое проигрывание проблемной ситуации;

анализ противоречивых высказываний - столкновение противоположных точек зрения на обсуждаемую проблему;

постановка проблемных вопросов;

альтернативный выбор (участникам предлагается выбрать одну из нескольких точек зрения или способов решения проблемы).

Методические рекомендации для студентов по отдельным формам самостоятельной работы

Система вузовского обучения подразумевает большую долю самостоятельности студентов в планировании и организации своей деятельности.

Работа с учебной литературой

При работе с учебной литературой необходимо подобрать литературу, научиться правильно ее читать, вести записи. Для подбора литературы в библиотеке используются алфавитный и систематический каталоги.

Правильный подбор учебной литературы рекомендуется преподавателем, читающим лекционный курс. Необходимая литература указана в методических разработках по данному курсу.

Изучая материал по выбранной литературе, следует переходить к следующему вопросу только после правильного уяснения предыдущего, описывая на бумаге все выкладки и вычисления (в том числе те, которые в учебнике опущены или на лекции даны для самостоятельного вывода).

Особое внимание следует обратить на определение основных понятий курса. Студент должен подробно разбирать примеры, которые поясняют такие определения, и уметь строить аналогичные примеры самостоятельно. Нужно добиваться точного представления о том, что изучаешь. Полезно составлять опорные конспекты. При изучении материала по учебнику полезно в тетради (на специально отведенных полях) дополнять конспект лекций. Там же следует отмечать вопросы, выделенные студентом для консультации с преподавателем.

Выводы, полученные в результате изучения, рекомендуется в конспекте выделять, чтобы они при перечитывании записей лучше запоминались.

Опыт показывает, что многим студентам помогает составление листа опорных сигналов, содержащего важнейшие и наиболее часто употребляемые понятия. Такой лист помогает запомнить основные положения лекции, а также может служить постоянным справочником для студента.

Различают два вида чтения: первичное и вторичное.

Первичное – это внимательное, неторопливое чтение, при котором можно остановиться на трудных местах. После него не должно остаться ни одного непонятого олова. Содержание не всегда может быть понятно после первичного чтения.

Задача *вторичного* чтения – полное усвоение смысла целого (по счету это чтение может быть и не вторым, а третьим или четвертым).

Правила самостоятельной работы с литературой

Выделяют **четыре основные установки в чтении учебно-научного текста**:

1. информационно-поисковый (задача – найти, выделить искомую информацию)
2. усваивающая (усилия читателя направлены на то, чтобы как можно полнее осознать и запомнить как сами сведения излагаемые автором, так и всю логику его рассуждений)
3. аналитико-критическая (читатель стремится критически осмыслить материал, проанализировав его, определив свое отношение к нему)
4. творческая (создает у читателя готовность в том или ином виде – как отправной пункт для своих рассуждений, как образ для действия по аналогии и т.п. – использовать суждения автора, ход его мыслей, результат наблюдения, разработанную методику, дополнить их, подвергнуть новой проверке).

С наличием различных установок обращения к научному тексту связано существование и нескольких **видов чтения**:

1. библиографическое – просматривание карточек каталога, рекомендательных списков, сводных списков журналов и статей за год и т.п.;
2. просмотрное – используется для поиска материалов, содержащих нужную информацию, обычно к нему прибегают сразу после работы со списками литературы и

каталогами, в результате такого просмотра читатель устанавливает, какие из источников будут использованы в дальнейшей работе;

3. ознакомительное – подразумевает сплошное, достаточно подробное прочтение отобранных статей, глав, отдельных страниц, цель – познакомиться с характером информации, узнать, какие вопросы вынесены автором на рассмотрение, провести сортировку материала;

4. изучающее – предполагает доскональное освоение материала; в ходе такого чтения проявляется доверие читателя к автору, готовность принять изложенную информацию, реализуется установка на предельно полное понимание материала;

5. аналитико-критическое и творческое чтение – два вида чтения близкие между собой тем, что участвуют в решении исследовательских задач. Первый из них предполагает направленный критический анализ, как самой информации, так и способов ее получения и подачи автором; второе – поиск тех суждений, фактов, по которым или в связи с которыми, читатель считает нужным высказать собственные мысли.

Из всех рассмотренных видов чтения основным для студентов является изучающее – именно оно позволяет в работе с учебной литературой накапливать знания в различных областях. Вот почему именно этот вид чтения в рамках учебной деятельности должен быть освоен в первую очередь. Кроме того, при овладении данным видом чтения формируются основные приемы, повышающие эффективность работы с научным текстом.

Основные виды систематизированной записи прочитанного:

1. **Аннотирование** – предельно краткое связное описание просмотренной или прочитанной книги (статьи), ее содержания, источников, характера и назначения;

2. **Планирование** – краткая логическая организация текста, раскрывающая содержание и структуру изучаемого материала;

3. **Тезирование** – лаконичное воспроизведение основных утверждений автора без привлечения фактического материала;

4. **Цитирование** – дословное выписывание из текста выдержек, извлечений, наиболее существенно отражающих ту или иную мысль автора;

5. **Конспектирование** – краткое и последовательное изложение содержания прочитанного.

Конспект – сложный способ изложения содержания книги или статьи в логической последовательности. Конспект аккумулирует в себе предыдущие виды записи, позволяет всесторонне охватить содержание книги, статьи. Умение составлять план, тезисы, делать выписки и другие записи определяет и технологию составления конспекта.

Методические рекомендации по составлению конспекта:

1. Внимательно прочитайте текст. Уточните в справочной литературе непонятные слова. При записи не забудьте вынести справочные данные на поля конспекта;

2. Выделите главное, составьте план;

3. Кратко сформулируйте основные положения текста, отметьте аргументацию автора;

4. Законспектируйте материал, четко следуя пунктам плана. При конспектировании старайтесь выразить мысль своими словами. Записи следует вести четко, ясно.

5. Грамотно записывайте цитаты. Цитируя, учитывайте лаконичность, значимость мысли.

В тексте конспекта желательно приводить не только тезисные положения, но и их доказательства. При оформлении конспекта необходимо стремиться к емкости каждого предложения. Мысли автора книги следует излагать кратко, заботясь о стиле и выразительности написанного. Число дополнительных элементов конспекта должно быть логически обоснованным, записи должны распределяться в определенной последовательности, отвечающей логической структуре произведения. Для уточнения и дополнения необходимо оставлять поля.

Овладение навыками конспектирования требует от студента целеустремленности, повседневной самостоятельной работы.

Самопроверка

После изучения определенной темы по записям в конспекте и учебнику, а также решения обсуждения проблемы на семинарских занятиях студенту рекомендуется, используя лист опорных сигналов, воспроизвести по памяти определения, выводы и формулировки основных положений и доказательств.

В случае необходимости нужно еще раз внимательно разобраться в материале.

Иногда недостаточность усвоения того или иного вопроса выясняется только при изучении дальнейшего материала. В этом случае надо вернуться назад и повторить плохо усвоенный материал. Важный критерий усвоения теоретического материала - пройти тестирование по пройденному материалу. Однако следует помнить, что правильное решение теста может получиться в результате применения механически заученных без понимания сущности теоретических положений.

Самопроверка включает:

- умение следить за собой: за своим поведением, речью, действиями и поступками, понимая при этом всю меру ответственности за них;
- умение контролировать степень понимания и степень прочности усвоения знаний и умений, познаваемых в учебном заведении, в коллективе, дома;
- умение критически оценивать результаты своей познавательной деятельности, в широком смысле – своих действий, поступков, труда (самооценка).

Самоконтроль учит ценить свое время, вырабатывает дисциплину труда (физического и умственного), позволяет вовремя заметить свои ошибки, вселяет веру в успешное использование знаний и умений на практике.

Способы самоконтроля могут быть следующими:

- перечитывание написанного текста и сравнение его с текстом учебной книги;
- повторное перечитывание материала с продумыванием его по частям;
- пересказ прочитанного;
- составление плана, тезисов, формулировок ключевых положений
- текста по памяти;
- рассказывание с опорой на иллюстрации, опорные положения;
- участие во взаимопроверке (анализ и оценка устных ответов, практических работ своих товарищей; дополнительные вопросы к их ответам; сочинения-рецензии и т.п.).

Самоконтроль является необходимым элементом учебного труда, прежде всего потому, что он способствует глубокому и прочному овладению знаниями.

Использование самоконтроля в учебной деятельности позволяет студенту оценивать эффективность и рациональность применяемых приемов и методов умственного труда, находить в нем допускаемые недочеты и на этой основе проводить необходимую его коррекцию.

И конечно, необходимо отметить большое воспитательное значение самоконтроля как оценочно-результативного компонента учебной деятельности. Овладение умениями самоконтроля приучает студентов к планированию учебного труда, способствует углублению их внимания, памяти и выступает как важный фактор развития познавательных способностей.

Текущие консультации

Если в процессе самостоятельной работы над изучением теоретического материала или при решении практических задач у студента возникают вопросы, разрешить которые самостоятельно не удастся, необходимо обратиться к преподавателю для получения у него разъяснений или указаний. В своих вопросах студент должен четко выразить, в чем он испытывает затруднения, характер этого затруднения. За консультацией следует обращаться и в случае, если возникнут сомнения в правильности ответов на вопросы самопроверки.

Методические рекомендации по подготовке к экзамену

Каждый учебный семестр заканчивается зачетно-экзаменационной сессией. Подготовка к зачетно-экзаменационной сессии, сдача зачетов и экзаменов является 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, слушал их невнимательно, 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.

Между экзаменами интервал 3-4 дня. Не следует думать, что 3-4 дня достаточно для успешной подготовки к экзаменам.

В эти 3-4 дня нужно систематизировать уже имеющиеся знания. На консультации перед экзаменом студентов познакомят с основными требованиями, ответят на возникшие у них вопросы. Поэтому посещение консультаций обязательно.

Требования к организации подготовки к экзаменам те же, что и при занятиях в течение семестра, но соблюдаться они должны более строго. Во-первых, очень важно соблюдение режима дня; сон не менее 8 часов в сутки, занятия заканчиваются не позднее, чем за 2-3 часа до сна. В перерывах между занятиями рекомендуются прогулки на свежем воздухе, неустойчивые занятия спортом. Во-вторых, наличие хороших собственных конспектов лекций. Даже в том случае, если была пропущена какая-либо лекция, необходимо во время ее восстановить (переписать ее на кафедре), обдумать, снять возникшие вопросы для того, чтобы запоминание материала было осознанным. В-третьих, при подготовке к экзаменам у студента должен быть хороший учебник или конспект литературы, прочитанной по указанию преподавателя в течение семестра. Здесь можно эффективно использовать листы опорных сигналов.

Вначале следует просмотреть весь материал по сдаваемой дисциплине, отметить для себя трудные вопросы. Обязательно в них разобраться. В заключение еще раз целесообразно повторить основные положения, используя при этом листы опорных сигналов.

Систематическая подготовка к занятиям в течение семестра позволит использовать время экзаменационной сессии для систематизации знаний.

9. ПЕРЕЧЕНЬ ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ.

При изучении дисциплины обучающимися используются следующие информационные технологии:

-аудиовизуальное представление обучающимся с помощью компьютера содержания отдельных тем дисциплины на лекционных занятиях;

-предоставление обучающимся доступа к учебному плану, рабочей программе дисциплины в электронной форме, к электронно-библиотечной системе института, содержащей учебно-методические материалы по дисциплине в электронной форме, к информационным справочным системам, которые используются при осуществлении образовательного процесса по дисциплине, посредством электронной информационно-образовательной среды института из любой точки, в которой имеется доступ к информационно-телекоммуникационной сети «Интернет»;

-фиксация хода образовательного процесса по дисциплине посредством электронной информационно-образовательной среды института;

-формирование электронного портфолио обучающегося по дисциплине посредством электронной информационно-образовательной среды института.

При осуществлении образовательного процесса по дисциплине используется следующее лицензионное программное обеспечение:

Word, Excel, PowerPoint;
Adobe Photoshop;
Adobe Premiere;
PowerDVD;
MediaPlayerClassic.

10. ОПИСАНИЕ МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЙ БАЗЫ, НЕОБХОДИМОЙ ДЛЯ ОСУЩЕСТВЛЕНИЯ ОБРАЗОВАТЕЛЬНОГО ПРОЦЕССА ПО ДИСЦИПЛИНЕ

Оборудованная учебная аудитория с маркерной доской, экраном и цифровым проектором для проведения лекционных и семинарских занятий. Библиотека с читальным залом; специализированная учебная лаборатория для практических занятий и самостоятельной работы бакалавров, оснащенная всем необходимым комплексом материально-технических средств, от копировальной техники, аудио-видеоустройств, до персональных компьютеров с выделенным доступом к сети Интернет.

11. Обеспечение образовательного процесса для лиц с ограниченными возможностями здоровья и инвалидов (при наличии)

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.

Составитель(и):

Ученая степень, звание, должность, Фамилия И.О.

.....

АННОТАЦИЯ ДИСЦИПЛИНЫ
Б1.В.12 ИНТЕЛЛЕКТУАЛЬНЫЙ АНАЛИЗ ДАННЫХ
09.03.02 Информационные системы и технологии
профиль/специализация
Информационные системы и цифровые технологии в культуре

Цель дисциплины (*модуля*):

Целью освоения дисциплины является: сформировать у студентов комплекс знаний и практических навыков в области интеллектуального анализа текстовых данных и дата майнинга, включая работу с большими коллекциями текстов, извлечение знаний и применение ИИ технологий.

Задачи:

- изучить методы обработки и анализа больших текстовых коллекций;
- освоить технологии дата-майнинга для извлечения знаний из текстов;
- научиться применять алгоритмы машинного обучения для классификации и кластеризации текстов;
- приобрести навыки тематического моделирования и сентимент-анализа больших корпусов;
- развить способность интерпретировать результаты анализа текстовых данных для решения прикладных задач.

Дисциплина (*модуль*) направлена на формирование следующих компетенций:

ПК-4

Готовность к информационно-аналитической деятельности и решению задач её автоматизации, интеллектуальному анализу данных

В результате освоения дисциплины (*модуля*) обучающийся должен:

Знать:

- основные методы обработки естественного языка (NLP) для больших данных;
- технологии дата майнинга текстовых коллекций;
- алгоритмы тематического моделирования (LDA, NMF) и кластеризации текстов;
- инструменты для работы с большими текстовыми корпусами (Spark NLP, Gensim);
- методы сентимент анализа и извлечения именованных сущностей (NER);
- принципы работы нейросетевых моделей для анализа текстов (BERT, GPT).

Уметь:

- проводить предобработку больших текстовых коллекций (токенизация, лемматизация, удаление стоп слов);
- применять методы дата майнинга для выявления закономерностей в текстах;
- строить тематические модели и интерпретировать их результаты;
- выполнять сентимент анализ отзывов и комментариев;
- визуализировать результаты анализа текстовых данных;
- использовать ИИ инструменты для автоматизации обработки текстов.

Владеть:

- навыками работы с библиотеками NLP (NLTK, SpaCy, Gensim, Transformers);
- методами тематического моделирования и кластеризации текстов;
- инструментами дата майнинга (Scikit learn, Spark NLP);
- приёмами работы с большими текстовыми корпусами и API ИИ сервисов.

По дисциплине (*модулю*) предусмотрена промежуточная аттестация в форме зачета и зачета с оценкой.

Общая трудоемкость освоения дисциплины (*модуля*) составляет 4 зачетных единиц.