

Документ подписан простой электронной подписью
Информация о владельце:
ФИО: Андрей Драгомирович Хлутков
Должность: директор
Дата подписания: 10.09.2026 23:18:48
Уникальный программный ключ:
880f7c07c583b07b775f6604a630281b13ca9fd2

**Федеральное государственное бюджетное образовательное
учреждение высшего образования
«РОССИЙСКАЯ АКАДЕМИЯ НАРОДНОГО ХОЗЯЙСТВА И ГОСУДАРСТВЕННОЙ
СЛУЖБЫ ПРИ ПРЕЗИДЕНТЕ РОССИЙСКОЙ ФЕДЕРАЦИИ»**

Северо-Западный институт управления - филиал РАНХиГС
«Факультет безопасности и таможи»
«Кафедра таможенного администрирования»

УТВЕРЖДЕНО

Директор

Северо-Западный института управления -
филиала РАНХиГС

Хлутков А.Д.

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ
Б1.В.ДВ.01.02 «Семантический анализ данных в профессиональной
деятельности»
САДвПД

по специальности: 38.05.02 «Таможенное дело»

краткое наименование дисциплины (модуля)

Специализация «Информационно-аналитическая деятельность в таможенном деле»

Квалификация: специалист таможенного дела

Формы обучения: очная

Год набора - 2022

Автор(ы)–составитель(и):

Кандидат технических наук, доцент кафедры таможенного администрирования, доцент
Г.Ю. Федотова

Доктор экономических наук, зав. кафедры таможенного администрирования, Гетман А.Г.

Заведующий кафедрой

таможенного администрирования

д-р.экон.н

А.Г. Гетман

РПД одобрена на заседании кафедры таможенного администрирования. Протокол от
(30.08.2022) № 1

С изменениями:

РПД одобрена на заседании кафедры таможенного администрирования. Протокол от (28.08.2026)
№ 1

СОДЕРЖАНИЕ

1. Перечень планируемых результатов обучения по дисциплине (модулю), соотнесенных с планируемыми результатами освоения образовательной программы
2. Объем и место дисциплины в структуре образовательной программы
3. Содержание и структура дисциплины
4. Материалы текущего контроля успеваемости обучающихся и фонд оценочных средств промежуточной аттестации по дисциплине
5. Методические указания для обучающихся по освоению дисциплины (модуля)
6. Учебная литература и ресурсы информационно-телекоммуникационной сети "Интернет", учебно-методическое обеспечение самостоятельной работы обучающихся по дисциплине
 - 6.1. Основная литература
 - 6.2. Дополнительная литература
 - 6.3. Учебно-методическое обеспечение самостоятельной работы
 - 6.4. Нормативные правовые документы
 - 6.5. Интернет-ресурсы
 - 6.6. Иные источники
7. Материально-техническая база, информационные технологии, программное обеспечение и информационные справочные системы

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

1.1. Дисциплина «Семантический анализ данных в профессиональной сфере» обеспечивает овладение следующими компетенциями:

Код компетенции	Наименование компетенции	Код этапа освоения компетенции	Наименование этапа освоения компетенции
ПК-37	владением методикой расчета показателей, отражающих результативность деятельности таможенных органов	ПК-37.2	Способность осуществлять семантический анализ данных в профессиональной деятельности

1.2. В результате освоения дисциплины у студентов должны быть сформированы:

ТФ	Код этапа освоения компетенции	Результаты обучения
Умение проводить анализ и прогнозировать поступление таможенных платежей в федеральный бюджет государства	ПК-37.2	на уровне знаний: принципов проведения семантического анализа по таможенным вопросам
		на уровне умений: применять методы семантического анализа в таможенной сфере
		на уровне навыков: проведения семантического анализа на отдельных этапах таможенного контроля

2. Объем и место дисциплины (модуля) в структуре ОП ВО

Объем дисциплины

Общая трудоемкость дисциплины составляет 3 зачетных единицы или 108 академических часов.

Для очной формы обучения трудоемкость контактной работы с преподавателем составляет 54 академических часа (из них 18 часов – лекции, 36 часов – практические занятия), самостоятельной работы – 54 академических часа.

Место дисциплины в структуре ОП ВО

Дисциплина «Семантический анализ данных в профессиональной сфере» относится к вариативной части учебного плана подготовки специалистов по специальности 38.05.02 «Таможенное дело».

Усвоение курса предполагает использование и интеграцию знаний и навыков, полученных студентам в ходе изучения большинства дисциплин.

Форма промежуточной аттестации в соответствии с учебным планом: зачет.

Дисциплина реализуется с применением дистанционных образовательных технологий

3. Содержание и структура дисциплины

Очная форма обучения

№ п/п	Наименование тем (разделов)	Всего	Объем дисциплины (модуля), час.						Форма текущего контроля успеваемости*, промежуточной аттестации
			Контактная работа обучающихся с преподавателем по видам учебных занятий				КС Р	СР	
			Л	ЛР	ПЗ	Конс .			
Тема 1.	Введение в анализ данных. Проблема обработки данных. Матрица данных.	12	2	0	4	0		6	Д,
Тема 2.	Назначение и этапы семантического анализа.	12	2	0	4	0		6	Т, РЗ
Тема 3.	Формализация в семантике.	13	2	0	4	0		7	Д
Тема 4.	Структурная семантика. Логическая семантика. Компьютерная семантика.	13	2	0	4	0		7	Т
Тема 5.	Методы интеллектуального анализа данных	13	2	0	4	0		7	УО
Тема 6.	Примеры практических приложений в таможенной сфере	13	2	0	4	0		7	Т
Тема 7.	Подготовка данных и интерпретация результатов в таможенной сфере	13	2	0	4	0		7	Д
Тема 8.	Практические аспекты применения семантического анализа данных в	19	4	0	8	0		7	Т

	таможенной сфере								
Промежуточная аттестация:									Зачет
Всего:		108	18	0	36	-		54	

Примечание: УО* – устный опрос; Т** – тестирование, Д - доклад
РЗ-решение задач

заочная форма

№ п/п	Наименование тем (разделов)	Всего	Объем дисциплины (модуля), час.					КС Р	СР	Форма текущего контроля успеваемости*, промежуточно й аттестации
			Контактная работа обучающихся с преподавателем по видам учебных занятий							
			Л	ЛР	ПЗ	Конс				
Тема 1.	Введение в анализ данных. Проблема обработки данных. Матрица данных.	12							12	Д,
Тема 2.	Назначение и этапы семантического анализа.	14	1		1				12	Т, РЗ
Тема 3.	Формализация в семантике.		1		1				12	Д
Тема 4.	Структурная семантика. Логическая семантика. Компьютерная семантика.	14	1		1				12	Т
Тема 5.	Методы интеллектуального анализа данных	14	1		1				12	УО
Тема 6.	Примеры практических приложений в таможенной сфере	13			1				12	Т
Тема 7.	Подготовка данных и интерпретация результатов в таможенной сфере	13			1				12	Д
Тема 8.	Практические аспекты применения семантического	10							10	Т

	анализа данных в таможенной сфере								
Промежуточная аттестация:		4						Зачет	
Всего:		108	4	0	6	-		94	

Содержание дисциплины

Тема 1: Введение в анализ данных. Проблема обработки данных. Матрица данных.

Принципы создания лингвистических теорий и моделей семантического уровня естественного языка. Связь между моделируемым объектом и его моделью. Структурные, логические, алгебраические, теоретико-множественные, топологические, статистические и другие формальные методы в семантических исследованиях.

Тема 2. Назначение и этапы семантического анализа.

Первичный семантический анализ. Алгоритм семантического анализа.

Тема 3. Формализация в семантике.

Применение классификаций и иерархий в семантике. Языковая и концептуальная картины мира, их роль в лингвистических описаниях. Фреймы, сценарии, семантические сети, логико-понятийные схемы и другие формы представления знаний. Понятие когнитивной структуры как способа организации знаний и их хранения в памяти. Семантический компонент в системах искусственного интеллекта. Компьютерная лексикография. Представление лексики в современных компьютерных тезаурусах. Лексические базы данных. Формальные онтологии.

Моделирование понимания текста и человеко-машинного взаимодействия. Семантика естественного языка в автоматических диалоговых системах, в системах машинного перевода и переводческой памяти. Семантическое обеспечение корпус-ориентированных исследований. Семантическая неоднозначность в естественном языке и методы ее разрешения. Автоматическое извлечение фактов из текстов, автоматическое индексирование и реферирование. Технологии Semantic Web.

Количественные методы в семантике. Проблемы лексикостатистики. Измерение количества семантической информации в тексте. Топологическая семантика и моделирование подмножеств лексикона с помощью метрических пространств. Меры сходства и расстояния. Кластерный анализ. Определение количественных оценок семантических связей единиц словаря и текста. Автоматическая классификация лексики и документов в корпусах текстов.

Тема 4. Структурная семантика. Логическая семантика. Компьютерная семантика.

Организация семантического пространства естественного языка в свете принципа композициональности значения. Компоненты значения языкового знака: семантические категории, семантические признаки, примитивы, прототипы и инварианты; их виды и внутренняя организация. Структура семантемы и комбинаторика семантических компонентов.

Логика как формальный язык описания структуры и значения языковых выражений. Модельно-теоретическая семантика. Типы естественных языковых выражений, их значение. Имена и дескрипции, их референциальные свойства. Пропозициональная функция. Предикаты, их логические и семантические типы. Логические операторы, кванторы, связки. Категория истинности. Принцип композициональности. Постулаты значения.

Тема 5. Методы интеллектуального анализа данных

Методы и инструменты создания, распространения и повторного использования семантических моделей. Методы и инструменты применения семантических моделей при формировании семантических хранилищ данных. Возможности и преимущества семантического анализа статистических данных. Методы и инструменты обогащения существующих массивов данных семантическими связями.

Семантические сети. Построение семантической сети из разнородных данных. Анализ текстовой информации. Технологии анализа данных: Data Mining, Visual Mining, Text Mining,

Тема 6. Примеры практических приложений в таможенной сфере.

Контроль описания товара. Контроль кода товара. Контроль таможенной стоимости товара. Контроль соблюдения запретов и ограничений участником ВЭД.

Тема 7. Подготовка данных и интерпретация результатов в таможенной сфере.

Документы и сведения в таможенной сфере как источник информации для применения методов семантического анализа.

Особенности подготовки данных для контроля достоверности заявленного кода.

Особенности подготовки данных для контроля таможенной стоимости товаров.

Особенности подготовки данных для контроля происхождения товаров.

Тема 8. Практические аспекты применения семантического анализа данных в таможенной сфере.

Применение семантического анализа на этапе предварительного информирования, проверки декларации на товары и таможенном контроле после выпуска товаров.

4. Материалы текущего контроля успеваемости обучающихся и фонд оценочных средств промежуточной аттестации по дисциплине

4.1. Формы и методы текущего контроля успеваемости обучающихся и промежуточной аттестации

В ходе реализации дисциплины «Семантический анализ данных в профессиональной сфере» используются следующие методы текущего контроля успеваемости обучающихся: : (в том числе, задания к контрольным точкам):

УО* – устный опрос; Т** – тестирование, Д - доклад

РЗ-решение задач

4. 2. Материалы текущего контроля успеваемости обучающихся

Полный перечень типовых оценочных материалов находится в ФОС по дисциплине

Тема 1.

Д,

Темы для докладов:

1. Методология Visual Mining: Зачем визуализировать семантические сети (графы), если у нас есть цифровые таблицы? Какую когнитивную проблему человека решает визуализация связей?

2. В чем разница между Data Mining (числа) и Text Mining (тексты) с точки зрения конечной цели? Что общего у них в математическом аппарате?

3. Опишите этапы построения семантической сети из «разнородных данных» (например, таблица Excel + новостная лента + геоданные). С чего начинается интеграция?

4. Какие методы разрешения сущностей (Entity Resolution) применяются при слиянии данных? (Как понять, что «А.С. Пушкин», «Пушкин Александр» и «Солнце русской поэзии» — это одна нода в сети?)

5. Что такое «инференция (вывод) по правилам» в семантической сети? Можно ли построить иерархию классификации автоматически, используя транзитивность («если А выше Б, а Б выше В, то А выше В»)?
6. Какие проблемы возникают при анализе текстовой информации на естественном языке для построения сети (неоднозначность, эллипсис, анафора)? Как они влияют на качество сети?
7. Критический вопрос: В Data Mining часто используются «черные ящики» (нейросети). Совместимы ли они с семантическими хранилищами, где важна прозрачность вывода (Explainable AI)? Приведите пример конфликта.
8. Сравните два подхода к кластеризации документов: на основе частотного вектора (TF-IDF) и на основе семантического расстояния в онтологии. Какой даст более осмысленные группы и почему?
9. Что такое «холодный старт» (Cold Start) для семантического хранилища? Как пополнить базу знаний семантическими связями, если корпус текстов маленький?
10. Как вы считаете, является ли построение семантической сети из разнородных данных задачей инженерии данных или лингвистической задачей? Аргументируйте свой ответ с точки зрения методов Data/Text Mining.

Тема 2.

Т,РЗ

Примеры тестов

1. Какой из методов наиболее подходит для автоматического сопоставления описания товара «концентрат яблочного сока» с кодом ТН ВЭД?
 - А) Кластеризация k-means
 - В) Семантический анализ с использованием словаря синонимов и правил
 - С) Линейная регрессия
 - Д) Метод главных компонент
2. Что означает термин «нормализация текста» в контексте семантического анализа таможенных деклараций?
 - А) Приведение текста к единому формату (регистр, сокращения, синонимы)
 - В) Удаление всех цифр из описания
 - С) Перевод текста на английский язык
 - Д) Разбиение текста на предложения
3. Для чего в таможенных системах используют словари синонимов?
 - А) Чтобы усложнить поиск товаров
 - В) Чтобы разные варианты названий одного товара считались эквивалентными
 - С) Чтобы увеличить объём базы данных
 - Д) Чтобы скрыть реальные наименования товаров

Примеры задач:

Задача 1 (Матрица данных)

Дан корпус из 3 предложений:

1. Я люблю кофе.
2. Ты любишь чай.
3. Он любит сок.

Постройте матрицу встречаемости слов (Bag of Words) без стоп-слов (исключите местоимения и союзы, оставьте глаголы и существительные). Укажите размерность матрицы.

Решение:

1. Токенизируем и удаляем местоимения («Я», «Ты», «Он» — их можно исключить, так как они не несут семантической нагрузки в данной модели либо оставить, но мы исключим по условию).
2. Оставляем: люблю, кофе, любишь, чай, любит, сок.
3. Приводим к леммам (начальная форма): любить, кофе, чай, сок.
4. Строим матрицу (Документы -> Слова):
 - D1: люблю (1), кофе (1), чай (0), сок (0)
 - D2: любишь (1), кофе (0), чай (1), сок (0)
 - D3: любит (1), кофе (0), чай (0), сок (1)Матрица (3 строки, 4 столбца):
[[1,1,0,0], [1,0,1,0], [1,0,0,1]]

Задача 2 (Принцип моделирования)

Объясните, почему логическая модель «Все люди смертны. Сократ — человек. => Сократ смертен» является гомоморфизмом реальности, а не изоморфизмом. Какой метод здесь использован?

Задача 3 (Теоретико-множественный метод)

Даны множества: $A = \{\text{слова со значением "перемещение по суше"}\}$ (идти, бежать, ползти), $B = \{\text{слова со значением "быстрое перемещение"}\}$ (бежать, лететь, мчаться). Найдите пересечение и объединение множеств. Что в семантике означает пересечение?

Задача 4 (Статистический метод)

В тексте из 1000 слов слово «свобода» встретилось 50 раз, а слово «закон» — 10 раз. Рассчитайте Term Frequency (TF) для этих слов. Можете ли вы на основе только TF утверждать, что текст о свободе? Почему проблематичен данный подход?

Задача 5 (Алгебраический метод)

Представьте семантику глагола "дать" (X дает Y Z) в виде формулы лямбда-исчисления. Опишите тип переменных.

Тема 3.

Д

Темы для докладов:

1. Формальные онтологии vs. Наивная языковая картина мира: конфликт или компромисс?
2. Семантические сети и фреймы: от «Ресторана» Минского до современных графов знаний (Knowledge Graphs).
3. Когнитивные структуры и память человека: как нейросети пытаются имитировать ассоциативное хранилище.
4. Семантический компонент в системах ИИ: почему «смысл» не сводится к «ответу»? Акцент: Проблема символов и тел (Symbol Grounding Problem). Как современные диалоговые системы (типа голосовых ассистентов) преобразуют звук в слоты фрейма и обратно в синтезированную речь. Ошибки в понимании дейксиса («положи это туда»).
5. Машинный перевод и переводческая память (ТМ): борьба статистики с семантикой. Акцент: Почему нейросетевой перевод (NMT) все равно «не понимает» текст, а лишь подбирает контекстные паттерны. Роль семантических ролей (кто кого?) при переводе с эргативных языков. Когда ТМ спасает переводчика, а когда — вредит (ложные друзья переводчика в контексте).

6. Автоматическое реферирование (Summarization) и извлечение фактов (NER): от данных к смыслу.
7. Семантическая неоднозначность (WSD) как главная головная боль NLP. Методы ее разрешения.
8. Лексикография 2.0: Компьютерные тезаурусы и лексические базы данных (WordNet, BabelNet, RuThes).
9. Semantic Web (Паутина смысла): стандарты RDF, OWL и микроразметка Schema.org.
10. Кластерный анализ и топологическая семантика: как измерить «смысловую близость» слов.

Тема 4.

Т

Примеры тестов:

1. Какой признак в описании товара может служить неформализованным индикатором риска при семантическом анализе?
 - A) Указание точного веса груза
 - B) Наличие слов «образец», «прототип», «для испытаний» при больших объёмах партии
 - C) Указание страны происхождения
 - D) Номер контейнера
2. В графе 31 декларации указано: «смесь для приготовления напитков, содержит мальтодекстрин, ароматизаторы, витамины». Какой семантический подход поможет определить, относится ли товар к категории детского питания?
 - A) Поиск ключевых слов «для детей», «детское» и анализ состава по нормативным перечням
 - B) Подсчёт количества слов в описании
 - C) Проверка длины строки
 - D) Анализ регистра букв
3. Что такое «стоп-слова» в семантическом анализе таможенных текстов?
 - A) Слова, которые всегда исключаются из анализа (предлоги, союзы и т.д.)
 - B) Слова, обозначающие опасные грузы
 - C) Слова, запрещённые к ввозу
 - D) Слова на иностранном языке

Тема 5.

УО

Вопросы:

1. Что такое «семантическая модель» в контексте хранилищ данных, и чем она принципиально отличается от реляционной модели (таблиц SQL)?
2. Перечислите основные стандарты (языки) для описания семантических моделей (например, RDF, RDFS, OWL). В чем разница в выразительной силе между RDFS и OWL?
3. В чем суть принципа «публикации связанных данных» (Linked Data)? Какие 5 звезд (5-Star Data) существуют для открытых данных?
4. Как инструменты создания онтологий (Protégé, TopBraid) позволяют обеспечить «повторное использование» модели? Что такое импорт (owl:imports) и зачем он нужен?
5. Что такое RDF-триплет (субъект-предикат-объект)? Приведите пример, как фраза «Санкт-Петербург основал Петр I» превращается в тройки.
6. В чем архитектурное преимущество графовых баз данных (семантических хранилищ) перед документарными NoSQL в задачах аналитики?

7. Что такое «обогащение данных семантическими связями» (Entity Linking)? Назовите главную проблему при связывании сущности из текста с сущностью в онтологии (омонимия/синонимия).
8. Объясните концепцию «Reasoning» (логического вывода) в семантическом хранилище. Если мы знаем, что «Собака — это Животное», и «Шарик — это Собака», сможет ли машина вывести, что «Шарик любит кости»? Почему да или почему нет?
9. Какие преимущества дает семантический анализ перед стандартным статистическим (числовым) анализом данных? Приведите пример, где цифры в столбце «Регион» ничего не скажут, а семантическая связь «столица/регион» — даст инсайт.
10. Что такое Text Mining, и на каком этапе обработки текста мы переходим от «кучи слов» к «семантическим данным»? (Вспомните про лемматизацию, распознавание именованных сущностей (NER) и синтаксический разбор).

тема 6.

Т

1 (Контроль описания товара)

При подаче декларации на товар «Куртка женская из натуральной кожи» система семантического контроля выявила несоответствие. В каком случае система должна выдать предупреждение?

- А) Если длина текста описания превышает 200 символов.
- В) Если в описании используется слово «кожа», но в графе «Материал верха» указан код полимерных материалов.
- С) Если цвет товара указан как «черный».
- D) Если страна происхождения товара совпадает со страной отправления.

2 (Контроль кода товара — ТН ВЭД)

В системе принято решение о классификации товара согласно ТН ВЭД. Какой метод ИИ наиболее адекватен для сопоставления текстового описания товара с кодом, если в справочнике есть иерархия (Разделы -> Группы -> Товарные позиции)?

- А) Простая линейная регрессия по длине слова.
- В) Применение многослойного персептрона без учета структуры иерархии.
- С) Построение семантической модели (онтологии) с учетом иерархических связей и синонимов материалов.
- D) Случайный выбор кода из 10 наиболее частотных вариантов.

3 (Контроль таможенной стоимости)

Для проверки корректности заявленной стоимости товара (инвойсной цены) система интеллектуального анализа данных сравнивает её с рыночными индикаторами. Какой метод используется для выявления аномалий в стоимости (андеррайтинг)?

- А) Метод кластеризации (поиск выбросов в группе однородных товаров).
- В) Пересчет валюты по текущему курсу ЦБ.
- С) Сравнение веса брутто и нетто.
- D) Проверка наличия всех реквизитов в контракте.

Вопрос 4 (Контроль соблюдения запретов и ограничений — лицензии, сертификаты)

Участник ВЭД подает декларацию на партию «Медицинское оборудование (рентгеновское)». Система семантического контроля обращается к нормативно-справочной информации (ЕАЭС). Что она должна проверить на первом этапе?

- А) Наличие разрешения на ввоз (лицензии), так как данная категория относится к подконтрольным товарам по решению Коллегии ЕЭК.
- В) Срок годности оборудования.
- С) Наличие печати организации-производителя на упаковке.

- D) Цвет корпуса аппарата.

5 (Комплексный — соответствие описания, кода и стоимости)

В чем заключается главное преимущество применения семантических сетей (онтологий) в таможенном контроле перед традиционными реляционными базами данных (SQL)?

- A) Семантические сети занимают меньше места на жестком диске.
- B) Они позволяют делать логические выводы (Reasoning) о противоречиях: например, если товар имеет цену \$500/кг, описан как «металлолом», но код ТН ВЭД требует кода «драгоценные металлы», система автоматически выявит нестыковку на основе цепочки свойств.
- C) Они быстрее выполняют SQL-запросы.
- D) Они не требуют участия эксперта в настройке.

тема 7.

Д,

Темы для докладов:

1. Документы ВЭД как неструктурированный источник: от скана контракта до RDF-триплета.
2. Проблема «мусорных» данных в таможенных декларациях: опечатки, аббревиатуры и нестандартные единицы измерения.
3. Иерархическая классификация vs. Плоская модель: почему нейросеть путает «насосы» и «компрессоры» при подготовке данных?
4. Роль Пояснений к ТН ВЭД и примечаний к разделам как источника семантических правил для подготовки данных.
5. Выявление ценовых аномалий: кластеризация однородных товаров и выбор бенчмарков (индикаторов).
6. Фактор инфляции и сезонности: как калибровать исторические данные о ценах перед сравнением.
7. Интерпретация полученных результатов (визуализация аномалий).
8. Контроль происхождения товаров: работа с сертификатами формы А, СТ-1 и критериями достаточной переработки.
9. Семантическая проверка цепочки поставок: как увязать данные инвойса, коносамента и сертификата о происхождении.
10. Интегративный дашборд (Dashboard) для таможенного инспектора: как не утонуть в трех видах аналитики (код, стоимость, страна).

тема-8

Т,

1 (Этап предварительного информирования)

При подаче предварительной информации о прибытии товаров морским транспортом система семантического анализа выявила, что в графе «отправитель» указана компания, зарегистрированная в офшорной зоне, а в графе «товар» — «оборудование двойного назначения». Какое действие должна выполнить интеллектуальная система на этом этапе?

- A) Отказать в регистрации предварительной информации.
- B) Присвоить высокий уровень риска и автоматически уведомить оперативный отдел о необходимости досмотра еще до прибытия судна.
- C) Самостоятельно скорректировать код ТН ВЭД на более безопасный.
- D) Проигнорировать, так как предварительная информация не имеет юридической силы.

2. (Проверка декларации на товары — корреляция признаков)

При проверке электронной декларации (ДТ) система семантического анализа данных обнаружила следующее противоречие:

- Вес брутто: 1 500 кг.
 - Вес нетто: 1 480 кг.
 - Код ТН ВЭД: 9403 (Мебель деревянная).
 - Страна происхождения: Бразилия.
 - Описание товара: «Детали для сборки мебели (ДСП)».
- Какой именно вид риска выявила система?
- А) Риск занижения таможенной стоимости.
 - В) Риск недостоверного описания (подмена материала: ДСП — это древесно-стружечная плита, а не массив дерева, что может менять код товара).
 - С) Риск нарушения валютного законодательства.
 - D) Риск неправильного подсчета веса (разница между брутто и нетто слишком мала для упаковки деревянной мебели).

3. (Таможенный контроль после выпуска товаров — камеральная проверка)

В рамках контроля после выпуска товаров (КПВТ) система анализирует массивы данных за прошедший год. С помощью каких методов интеллектуального анализа можно выявить фирмы-однодневки, использующие схемы «лже-экспорта»?

- А) Только визуальный просмотр каждой декларации инспектором.
- В) Применение алгоритмов кластеризации и поиска аномалий (обнаружение сообществ в графе поставок: если группа компаний замыкает цепочку платежей сама на себя).
- С) Проверка суммы НДС в каждой отдельной декларации.
- D) Сравнение подписей на бумажных носителях.

4. (Семантический анализ текста в ДТ — распознавание целей)

Декларант подал ДТ с описанием товара: «Части оборудования для дробления камня, бывшие в употреблении, для строительной отрасли». Система семантического анализа, используя онтологию, находит скрытый признак. Какой признак проверяется в первую очередь при наличии слов «бывшие в употреблении»?

- А) Цвет товара.
- В) Необходимость представления разрешения Росприроднадзора (проверка на отходы/металлолом).
- С) Проверка курса валюты.
- D) Соответствие размера упаковки.

5. (Интеграция трех этапов в единую систему)

В чем заключается ключевое отличие применения семантического анализа на этапе после выпуска от этапа предварительного информирования с точки зрения подготовки данных?

- А) На этапе после выпуска данные не требуют нормализации, так как они уже проверены.
- В) На этапе предварительного информирования данных мало, а на этапе после выпуска система оперирует историческими массивами (Big Data) и может строить сложные корреляции (например, связь курса валюты и цены контракта).
- С) На этапе после выпуска семантика вообще не применяется, только арифметика.
- D) Отличий нет, используется один и тот же алгоритм.

Для каждой формы текущего контроля успеваемости обучающихся в рамках КТ определены критерии оценивания результатов выполнения задания.

Общая шкала оценивания результатов текущего контроля успеваемости и промежуточной аттестации обучающихся с применением БРС

Итоговая балльная оценка	Традиционная система	Бинарная система	ECTS	
			Для традиционной системы	Для бинарной системы
95-100	Отлично	Зачтено	A	P/ Passed
85-94			B	P/ Passed
75-84	Хорошо		C	P/ Passed
65-74			D	P/ Passed
55-64	Удовлетворительно		E	P/ Passed
0-54	Неудовлетворительно	Не зачтено	F	F/Failed

Соотношение баллов за текущий контроль успеваемости и промежуточную аттестацию, а также повторную промежуточную аттестацию:

Максимальная сумма баллов за текущий контроль успеваемости	Максимальная сумма баллов за промежуточную аттестацию	Максимальная итоговая балльная оценка	Максимальная сумма баллов за повторную промежуточную аттестацию
60 баллов	40 баллов	100 баллов	100 баллов

Оценочные материалы промежуточной аттестации по дисциплине

Один или несколько тематических блоков дисциплины завершаются контрольной точкой (далее – КТ). Текущий контроль успеваемости по дисциплине предусматривает не менее 2 (двух) и не более 10 (десяти) КТ в течение периода освоения дисциплины.

Максимальное количество баллов за любой тип работ в рамках КТ составляет 100 (сто) баллов.

Распределение весовых коэффициентов по КТ в рамках текущего контроля успеваемости по дисциплине и формулы расчета:

Наименование контрольной точки	Максимальное количество баллов за работу в рамках КТ, которое может набрать студент	Коэффициент веса контрольной точки	Результат контрольной точки, участвующий в формировании итоговой балльной оценки по дисциплине (отражается в журнале БРС в СДО)
КТ 1	100	0,1	10
КТ 2	100	0,1	10
КТ 3	100	0,1	10
КТ 4	100	0,1	10
КТ 5	100	0,05	5
КТ 6	100	0,05	5
КТ 7	100	0,05	5
КТ 8	100	0,05	5
Итого:	х	0,6	60

Формула расчета результата контрольной точки:

Результат контрольной точки = Количество баллов за работу в рамках КТ X Коэффициент веса контрольной точки.

Формы текущего контроля успеваемости обучающихся в рамках КТ и типовые оценочные материалы:

КТ-1

Тема 1.

Д,

Темы для докладов:

1. Методология Visual Mining: Зачем визуализировать семантические сети (графы), если у нас есть цифровые таблицы? Какую когнитивную проблему человека решает визуализация связей?
2. В чем разница между Data Mining (числа) и Text Mining (тексты) с точки зрения конечной цели? Что общего у них в математическом аппарате?
3. Опишите этапы построения семантической сети из «разнородных данных» (например, таблица Excel + новостная лента + геоданные). С чего начинается интеграция?
4. Какие методы разрешения сущностей (Entity Resolution) применяются при слиянии данных? (Как понять, что «А.С. Пушкин», «Пушкин Александр» и «Солнце русской поэзии» — это одна нода в сети?)
5. Что такое «инференция (вывод) по правилам» в семантической сети? Можно ли построить иерархию классификации автоматически, используя транзитивность («если А выше Б, а Б выше В, то А выше В»)?
6. Какие проблемы возникают при анализе текстовой информации на естественном языке для построения сети (неоднозначность, эллипсис, анафора)? Как они влияют на качество сети?
7. Критический вопрос: В Data Mining часто используются «черные ящики» (нейросети). Совместимы ли они с семантическими хранилищами, где важна прозрачность вывода (Explainable AI)? Приведите пример конфликта.

8. Сравните два подхода к кластеризации документов: на основе частотного вектора (TF-IDF) и на основе семантического расстояния в онтологии. Какой даст более осмысленные группы и почему?
9. Что такое «холодный старт» (Cold Start) для семантического хранилища? Как пополнить базу знаний семантическими связями, если корпус текстов маленький?
10. Как вы считаете, является ли построение семантической сети из разнородных данных задачей инженерии данных или лингвистической задачей? Аргументируйте свой ответ с точки зрения методов Data/Text Mining.

КТ-2

Тема 2.

ТРЗ

Примеры тестов

1. Какой из методов наиболее подходит для автоматического сопоставления описания товара «концентрат яблочного сока» с кодом ТН ВЭД?
- А) Кластеризация k-means
 - В) Семантический анализ с использованием словаря синонимов и правил
 - С) Линейная регрессия
 - Д) Метод главных компонент
2. Что означает термин «нормализация текста» в контексте семантического анализа таможенных деклараций?
- А) Приведение текста к единому формату (регистр, сокращения, синонимы)
 - В) Удаление всех цифр из описания
 - С) Перевод текста на английский язык
 - Д) Разбиение текста на предложения
3. Для чего в таможенных системах используют словари синонимов?
- А) Чтобы усложнить поиск товаров
 - В) Чтобы разные варианты названий одного товара считались эквивалентными
 - С) Чтобы увеличить объём базы данных
 - Д) Чтобы скрыть реальные наименования товаров

Примеры задач:

Задача 1 (Матрица данных)

Дан корпус из 3 предложений:

1. Я люблю кофе.
2. Ты любишь чай.
3. Он любит сок.

Постройте матрицу встречаемости слов (Bag of Words) без стоп-слов (исключите местоимения и союзы, оставьте глаголы и существительные). Укажите размерность матрицы.

Задача 2 (Принцип моделирования)

Объясните, почему логическая модель «Все люди смертны. Сократ — человек. => Сократ смертен» является гомоморфизмом реальности, а не изоморфизмом. Какой метод здесь использован?

Задача 3 (Теоретико-множественный метод)

Даны множества: $A = \{\text{слова со значением "перемещение по суше"}\}$ (идти, бежать, ползти), $B = \{\text{слова со значением "быстрое перемещение"}\}$ (бежать, лететь, мчаться). Найдите пересечение и объединение множеств. Что в семантике означает пересечение?

Задача 4 (Статистический метод)

В тексте из 1000 слов слово «свобода» встретилось 50 раз, а слово «закон» — 10 раз. Рассчитайте Term Frequency (TF) для этих слов. Можете ли вы на основе только TF утверждать, что текст о свободе? Почему проблематичен данный подход?

Задача 5 (Алгебраический метод)

Представьте семантику глагола "дать" (X дает Y Z) в виде формулы лямбда-исчисления. Опишите тип переменных.

КТ-3

Тема 3.

Д

Темы для докладов:

1. Формальные онтологии vs. Наивная языковая картина мира: конфликт или компромисс?
2. Семантические сети и фреймы: от «Ресторана» Минского до современных графов знаний (Knowledge Graphs).
3. Когнитивные структуры и память человека: как нейросети пытаются имитировать ассоциативное хранилище.
4. Семантический компонент в системах ИИ: почему «смысл» не сводится к «ответу»? Акцент: Проблема символов и тел (Symbol Grounding Problem). Как современные диалоговые системы (типа голосовых ассистентов) преобразуют звук в слоты фрейма и обратно в синтезированную речь. Ошибки в понимании дейксиса («положи это туда»).
5. Машинный перевод и переводческая память (ТМ): борьба статистики с семантикой. Акцент: Почему нейросетевой перевод (NMT) все равно «не понимает» текст, а лишь подбирает контекстные паттерны. Роль семантических ролей (кто кого?) при переводе с эргативных языков. Когда ТМ спасает переводчика, а когда — вредит (ложные друзья переводчика в контексте).
6. Автоматическое реферирование (Summarization) и извлечение фактов (NER): от данных к смыслу.
7. Семантическая неоднозначность (WSD) как главная головная боль NLP. Методы ее разрешения.
8. Лексикография 2.0: Компьютерные тезаурусы и лексические базы данных (WordNet, BabelNet, RuThes).
9. Semantic Web (Паутина смысла): стандарты RDF, OWL и микроразметка Schema.org.
10. Кластерный анализ и топологическая семантика: как измерить «смысловую близость» слов.

КТ-4

Тема 4.

Т

Примеры тестов:

1. Какой признак в описании товара может служить неформализованным индикатором риска при семантическом анализе?
А) Указание точного веса груза
В) Наличие слов «образец», «прототип», «для испытаний» при больших объемах партии

- C) Указание страны происхождения
- D) Номер контейнера

2. В графе 31 декларации указано: «смесь для приготовления напитков, содержит мальтодекстрин, ароматизаторы, витамины». Какой семантический подход поможет определить, относится ли товар к категории детского питания?

- A) Поиск ключевых слов «для детей», «детское» и анализ состава по нормативным перечням
- B) Подсчёт количества слов в описании
- C) Проверка длины строки
- D) Анализ регистра букв

3. Что такое «стоп-слова» в семантическом анализе таможенных текстов?

- A) Слова, которые всегда исключаются из анализа (предлоги, союзы и т.д.)
- B) Слова, обозначающие опасные грузы
- C) Слова, запрещённые к ввозу
- D) Слова на иностранном языке

КТ-5

Тема 5.

УО

Вопросы:

1. Что такое «семантическая модель» в контексте хранилищ данных, и чем она принципиально отличается от реляционной модели (таблиц SQL)?
2. Перечислите основные стандарты (языки) для описания семантических моделей (например, RDF, RDFS, OWL). В чем разница в выразительной силе между RDFS и OWL?
3. В чем суть принципа «публикации связанных данных» (Linked Data)? Какие 5 звезд (5-Star Data) существуют для открытых данных?
4. Как инструменты создания онтологий (Protégé, TopBraid) позволяют обеспечить «повторное использование» модели? Что такое импорт (owl:imports) и зачем он нужен?
5. Что такое RDF-триплет (субъект-предикат-объект)? Приведите пример, как фраза «Санкт-Петербург основал Петр I» превращается в тройки.
6. В чем архитектурное преимущество графовых баз данных (семантических хранилищ) перед документарными NoSQL в задачах аналитики?
7. Что такое «обогащение данных семантическими связями» (Entity Linking)? Назовите главную проблему при связывании сущности из текста с сущностью в онтологии (омонимия/синонимия).
8. Объясните концепцию «Reasoning» (логического вывода) в семантическом хранилище. Если мы знаем, что «Собака — это Животное», и «Шарик — это Собака», сможет ли машина вывести, что «Шарик любит кости»? Почему да или почему нет?
9. Какие преимущества дает семантический анализ перед стандартным статистическим (числовым) анализом данных? Приведите пример, где цифры в столбце «Регион» ничего не скажут, а семантическая связь «столица/регион» — даст инсайт.
10. Что такое Text Mining, и на каком этапе обработки текста мы переходим от «кучи слов» к «семантическим данным»? (Вспомните про лемматизацию, распознавание именованных сущностей (NER) и синтаксический разбор).

КТ-6

тема 6.

Т

1 (Контроль описания товара)

При подаче декларации на товар «Куртка женская из натуральной кожи» система семантического контроля выявила несоответствие. В каком случае система должна выдать предупреждение?

- А) Если длина текста описания превышает 200 символов.
- В) Если в описании используется слово «кожа», но в графе «Материал верха» указан код полимерных материалов.
- С) Если цвет товара указан как «черный».
- D) Если страна происхождения товара совпадает со страной отправления.

2 (Контроль кода товара — ТН ВЭД)

В системе принято решение о классификации товара согласно ТН ВЭД. Какой метод ИИ наиболее адекватен для сопоставления текстового описания товара с кодом, если в справочнике есть иерархия (Разделы -> Группы -> Товарные позиции)?

- А) Простая линейная регрессия по длине слова.
- В) Применение многослойного персептрона без учета структуры иерархии.
- С) Построение семантической модели (онтологии) с учетом иерархических связей и синонимов материалов.
- D) Случайный выбор кода из 10 наиболее частотных вариантов.

3 (Контроль таможенной стоимости)

Для проверки корректности заявленной стоимости товара (инвойсной цены) система интеллектуального анализа данных сравнивает её с рыночными индикаторами. Какой метод используется для выявления аномалий в стоимости (андеррайтинг)?

- А) Метод кластеризации (поиск выбросов в группе однородных товаров).
- В) Пересчет валюты по текущему курсу ЦБ.
- С) Сравнение веса брутто и нетто.
- D) Проверка наличия всех реквизитов в контракте.

Вопрос 4 (Контроль соблюдения запретов и ограничений — лицензии, сертификаты)

Участник ВЭД подает декларацию на партию «Медицинское оборудование (рентгеновское)». Система семантического контроля обращается к нормативно-справочной информации (ЕАЭС). Что она должна проверить на первом этапе?

- А) Наличие разрешения на ввоз (лицензии), так как данная категория относится к подконтрольным товарам по решению Коллегии ЕЭК.
- В) Срок годности оборудования.
- С) Наличие печати организации-производителя на упаковке.
- D) Цвет корпуса аппарата.

5 (Комплексный — соответствие описания, кода и стоимости)

В чем заключается главное преимущество применения семантических сетей (онтологий) в таможенном контроле перед традиционными реляционными базами данных (SQL)?

- А) Семантические сети занимают меньше места на жестком диске.
- В) Они позволяют делать логические выводы (Reasoning) о противоречиях: например, если товар имеет цену \$500/кг, описан как «металлолом», но код ТН ВЭД требует кода «драгоценные металлы», система автоматически выявит нестыковку на основе цепочки свойств.
- С) Они быстрее выполняют SQL-запросы.
- D) Они не требуют участия эксперта в настройке.

КТ-7

тема 7.

Д,

Темы для докладов:

1. Документы ВЭД как неструктурированный источник: от скана контракта до RDF-триплета.
2. Проблема «мусорных» данных в таможенных декларациях: опечатки, аббревиатуры и нестандартные единицы измерения.
3. Иерархическая классификация vs. Плоская модель: почему нейросеть путает «насосы» и «компрессоры» при подготовке данных?
4. Роль Пояснений к ТН ВЭД и примечаний к разделам как источника семантических правил для подготовки данных.
5. Выявление ценовых аномалий: кластеризация однородных товаров и выбор бенчмарков (индикаторов).
6. Фактор инфляции и сезонности: как калибровать исторические данные о ценах перед сравнением.
7. Интерпретация полученных результатов (визуализация аномалий).
8. Контроль происхождения товаров: работа с сертификатами формы А, СТ-1 и критериями достаточной переработки.
9. Семантическая проверка цепочки поставок: как увязать данные инвойса, коносамента и сертификата о происхождении.
10. Интегративный дашборд (Dashboard) для таможенного инспектора: как не утонуть в трех видах аналитики (код, стоимость, страна).

КТ-8

тема-8

Т,

1 (Этап предварительного информирования)

При подаче предварительной информации о прибытии товаров морским транспортом система семантического анализа выявила, что в графе «отправитель» указана компания, зарегистрированная в офшорной зоне, а в графе «товар» — «оборудование двойного назначения». Какое действие должна выполнить интеллектуальная система на этом этапе?

- А) Отказать в регистрации предварительной информации.
- В) Присвоить высокий уровень риска и автоматически уведомить оперативный отдел о необходимости досмотра еще до прибытия судна.
- С) Самостоятельно скорректировать код ТН ВЭД на более безопасный.
- D) Проигнорировать, так как предварительная информация не имеет юридической силы.

2. (Проверка декларации на товары — корреляция признаков)

При проверке электронной декларации (ДТ) система семантического анализа данных обнаружила следующее противоречие:

- Вес брутто: 1 500 кг.
- Вес нетто: 1 480 кг.
- Код ТН ВЭД: 9403 (Мебель деревянная).
- Страна происхождения: Бразилия.
- Описание товара: «Детали для сборки мебели (ДСП)».

Какой именно вид риска выявила система?

- А) Риск занижения таможенной стоимости.
- В) Риск недостоверного описания (подмена материала: ДСП — это древесно-стружечная плита, а не массив дерева, что может менять код товара).
- С) Риск нарушения валютного законодательства.

- D) Риск неправильного подсчета веса (разница между брутто и нетто слишком мала для упаковки деревянной мебели).

3. (Таможенный контроль после выпуска товаров — камеральная проверка)

В рамках контроля после выпуска товаров (КПВТ) система анализирует массивы данных за прошедший год. С помощью каких методов интеллектуального анализа можно выявить фирмы-однодневки, использующие схемы «лже-экспорта»?

- A) Только визуальный просмотр каждой декларации инспектором.
- B) Применение алгоритмов кластеризации и поиска аномалий (обнаружение сообществ в графе поставок: если группа компаний замыкает цепочку платежей сама на себя).
- C) Проверка суммы НДС в каждой отдельной декларации.
- D) Сравнение подписей на бумажных носителях.

4. (Семантический анализ текста в ДТ — распознавание целей)

Декларант подал ДТ с описанием товара: «Части оборудования для дробления камня, бывшие в употреблении, для строительной отрасли». Система семантического анализа, используя онтологию, находит скрытый признак. Какой признак проверяется в первую очередь при наличии слов «бывшие в употреблении»?

- A) Цвет товара.
- B) Необходимость представления разрешения Росприроднадзора (проверка на отходы/металлолом).
- C) Проверка курса валюты.
- D) Соответствие размера упаковки.

5. (Интеграция трех этапов в единую систему)

В чем заключается ключевое отличие применения семантического анализа на этапе после выпуска от этапа предварительного информирования с точки зрения подготовки данных?

- A) На этапе после выпуска данные не требуют нормализации, так как они уже проверены.
- B) На этапе предварительного информирования данных мало, а на этапе после выпуска система оперирует историческими массивами (Big Data) и может строить сложные корреляции (например, связь курса валюты и цены контракта).
- C) На этапе после выпуска семантика вообще не применяется, только арифметика.
- D) Отличий нет, используется один и тот же алгоритм.

Для каждой формы текущего контроля успеваемости обучающихся в рамках КТ определены критерии оценивания результатов выполнения задания.

Критерии оценивания тестирования (Т):

Критерии оценки	Диапазон баллов	Описание критерия
Количество правильных ответов	0	Количество правильных ответов менее 55%
	25	Количество правильных ответов от 55% до 64%
	50	Количество правильных ответов от 65% до 74%
	75	Количество правильных ответов от 75% до 84%
	100	Количество правильных ответов от 85% до 100%
Итого максимально:	100	

Критерии оценивания устного опроса (УО):

Критерии оценки	Диапазон баллов	Описание критерия
-----------------	-----------------	-------------------

Качество правильных ответов	0-54	Обучающийся обнаруживает незнание вопроса, допускает ошибки в формулировке определений и правил, искажающие их смысл, беспорядочно и неуверенно излагает материал.
	55-64	Обучающийся обнаруживает знание и понимание основных положений данной темы, но излагает материал неполно и допускает неточности в определении понятий или формулировке правил; не умеет достаточно глубоко и доказательно обосновать свои суждения и привести свои примеры; излагает материал непоследовательно и допускает ошибки в языковом оформлении излагаемого.
	65-84	Обучающийся дает ответ, удовлетворяющий тем же требованиям, что и для оценки «отлично», но допускает 1–2 ошибки, которые сам же исправляет, и 1–2 недочета в последовательности и языковом оформлении излагаемого.
	85-100	Обучающийся полно излагает материал (отвечает на вопрос), дает правильное определение основных понятий; обнаруживает понимание материала, может обосновать свои суждения, применить знания на практике, привести необходимые примеры не только из учебника, но и самостоятельно составленные; излагает материал последовательно и правильно с точки зрения норм литературного языка.
Итого максимально:		100

Критерии оценивания расчетных заданий (РЗ):

Критерии оценки	Диапазон баллов	Описание критерия
Качество правильных ответов	0-54	У обучающегося отсутствуют ответы на большинство вопросов задачи, задание не выполнено или выполнено не верно.
	55-64	Обучающимся допущены отдельные ошибки при выполнении задания
	65-84	Обучающимся в целом задание выполнено, имеются отдельные неточности или недостаточно полные ответы, не содержащие ошибок.
	85-100	Обучающимся задание выполнено без ошибок и в полном объеме.
Итого максимально:		100

Критерии оценивания доклада (Д)

Критерии оценки	Диапазон баллов	Описание критерия
Содержание и раскрытие темы	0-20	Детальное, последовательное описание всех этапов с конкретными примерами
Грамотность изложения	0-20	Соблюдены все правила грамматики, орфографии и пунктуации
Стилистика	0-20	Единый стиль изложения, точные формулировки, уместное использование терминов, лаконичность
Логика изложения	0-20	Чёткая последовательность изложения, логические связи между частями текста, аргументы подтверждают выводы
Оригинальность	0-20	Уникальный подход к теме, нестандартные решения, инновационные идеи, собственная позиция автора
Итого максимально:	100	

Описание дополнительных материалов и оборудования, необходимых для выполнения проверочных заданий.

Для решения ситуационных заданий, тестовых заданий студенту разрешается использование калькулятора; программ для работы с электронными таблицами для обработки, анализа и визуализации данных.

Формы промежуточной аттестации, критерии и шкала оценивания, типовые оценочные материалы по дисциплине

Промежуточная аттестация проводится в форме зачета
может проводиться с применением следующих методов (средств):

устный ответ на вопросы билета;
письменный ответ на вопросы билета.

В каждом билете не менее 2 вопросов.

При реализации промежуточной аттестации с применением ДОТ:

устно с прокторингом – в форме ответа на вопросы билета;
письменно с прокторингом – в форме ответа на вопросы билета.

Описание дополнительных материалов и оборудования, необходимых для выполнения проверочных заданий (при необходимости).

Для решения ситуационных заданий, тестовых заданий студенту разрешается использование калькулятора; программ для работы с электронными таблицами для обработки, анализа и визуализации данных.

Вопросы для подготовки к зачету:

1. Что такое объекты и признаки в семантическом анализе? Для чего нужен функционал качества? Что такое алгоритм/модель? Что такое вещественные (числовые), бинарные, категориальные признаки? Приведите примеры.
2. В чём заключаются задачи классификации, регрессии и кластеризации? Приведите примеры для каждого типа задач.
3. В чём заключается обобщающая способность алгоритма семантического анализа?
4. Организация логического вывода.

5. Суть метода резолюции.
6. Проблемы логического вывода.
7. Организация вывода в семантических сетях и фреймовых моделях.
8. Принципы создания лингвистических теорий и моделей семантического уровня естественного языка. Связь между моделируемым объектом и его моделью. Структурные, логические, алгебраические, теоретико-множественные, топологические, статистические и другие формальные методы в семантических исследованиях.
9. Первичный семантический анализ.
10. Алгоритм семантического анализа.
11. Применение классификаций и иерархий в семантике. Языковая и концептуальная картины мира, их роль в лингвистических описаниях. Фреймы, сценарии, семантические сети, логико-понятийные схемы и другие формы представления знаний.
12. Понятие когнитивной структуры как способа организации знаний и их хранения в памяти.
13. Семантический компонент в системах искусственного интеллекта.
14. Компьютерная лексикография. Представление лексики в современных компьютерных тезаурусах. Лексические базы данных. Формальные онтологии.
15. Моделирование понимания текста и человеко-машинного взаимодействия. Семантика естественного языка в автоматических диалоговых системах, в системах машинного перевода и переводческой памяти.
16. Семантическое обеспечение корпус-ориентированных исследований. Семантическая неоднозначность в естественном языке и методы ее разрешения.
17. Автоматическое извлечение фактов из текстов, автоматическое индексирование и реферирование. Технологии Semantic Web.
18. Количественные методы в семантике. Проблемы лексикостатистики.
19. Измерение количества семантической информации в тексте. Топологическая семантика и моделирование подмножеств лексикона с помощью метрических пространств. Меры сходства и расстояния.
20. Кластерный анализ. Определение количественных оценок семантических связей единиц словаря и текста.
21. Автоматическая классификация лексики и документов в корпусах текстов.
22. Организация семантического пространства естественного языка в свете принципа композиционности значения. Компоненты значения языкового знака: семантические категории, семантические признаки, примитивы, прототипы и инварианты; их виды и внутренняя организация. Структура семантемы и комбинаторика семантических компонентов.
23. Логика как формальный язык описания структуры и значения языковых выражений. Модельно-теоретическая семантика.
24. Типы естественных языковых выражений, их значение.
25. Имена и дескрипции, их референциальные свойства. Пропозициональная функция.
26. Предикаты, их логические и семантические типы. Логические операторы, кванторы, связи. Категория истинности. Принцип композициональности. Постулаты значения.
27. Методы и инструменты создания, распространения и повторного использования семантических моделей.

28. Методы и инструменты применения семантических моделей при формировании семантических хранилищ данных. Возможности и преимущества семантического анализа статистических данных.
29. Методы и инструменты обогащения существующих массивов данных семантическими связями.
30. Семантические сети.
31. Построение семантической сети из разнородных данных. Анализ текстовой информации.
32. Технологии анализа данных: Data Mining, Visual Mining, Text Mining,
33. Контроль описания товара на основе применения семантического анализа данных.
34. Контроль кода товара на основе применения семантического анализа данных.
35. Контроль таможенной стоимости товара на основе применения семантического анализа данных.
36. Контроль соблюдения запретов и ограничений участником ВЭД на основе применения семантического анализа данных.
37. Документы и сведения в таможенной сфере как источник информации для применения методов семантического анализа.
38. Особенности подготовки данных для контроля достоверности заявленного кода.
39. Особенности подготовки данных для контроля таможенной стоимости товаров.
40. Особенности подготовки данных для контроля происхождения товаров.
41. Применение семантического анализа на этапе предварительного информирования.
42. Применение семантического анализа на этапе проверки декларации на товары.
43. Применение семантического анализа при таможенном контроле после выпуска товаров.

Критерии и шкала оценивания на основе БРС.

Критерии и балльная шкала определяются преподавателем

КРИТЕРИИ ОЦЕНИВАНИЯ	РЕЗУЛЬТАТ В БАЛЛАХ
Дан полный, в логической последовательности развернутый ответ на поставленный вопрос, где он продемонстрировал знания предмета в полном объеме учебной программы, достаточно глубоко осмысливает дисциплину, самостоятельно, и исчерпывающе отвечает на дополнительные вопросы, приводит собственные примеры по проблематике поставленного вопроса, решил предложенные практические задания без ошибок	40
Дан развернутый ответ на поставленный вопрос, где студент демонстрирует знания, приобретенные на лекционных и семинарских занятиях, а также полученные посредством изучения обязательных учебных материалов по курсу, дает аргументированные ответы, приводит примеры, в ответе присутствует свободное владение монологической речью, логичность и последовательность ответа. Однако допускается неточность в ответе. Решил предложенные практические задания с небольшими неточностями.	30-39
Дан ответ, свидетельствующий в основном о знании процессов изучаемой дисциплины, отличающийся недостаточной глубиной и полнотой раскрытия темы, знанием основных вопросов теории, слабо сформированными навыками анализа	20-29

явлений, процессов, недостаточным умением давать аргументированные ответы и приводить примеры, недостаточно свободным владением монологической речью, логичностью и последовательностью ответа. Допускается несколько ошибок в содержании ответа и решении практических заданий.	
Дан ответ, который содержит ряд серьезных неточностей, обнаруживающий незнание процессов изучаемой предметной области, отличающийся неглубоким раскрытием темы, незнанием основных вопросов теории, несформированными навыками анализа явлений, процессов, неумением давать аргументированные ответы, слабым владением монологической речью, отсутствием логичности и последовательности. Выводы поверхностны. Решение практических заданий не выполнено, т.е. студент не способен ответить на вопросы даже при дополнительных наводящих вопросах преподавателя.	0-19

Описание дополнительных материалов и оборудования, необходимых для выполнения проверочных заданий (при необходимости).

Для решения ситуационных заданий, тестовых заданий студенту разрешается использование калькулятора; программ для работы с электронными таблицами для обработки, анализа и визуализации данных.

4.4. Методические материалы

Дисциплина «Семантический анализ данных в профессиональной сфере», изучается студентами в седьмом (шестом и седьмом для заочного обучения) семестре. При подготовке к лекционным занятиям студенту следует ознакомиться с учебно-тематическим планом изучаемой учебной дисциплины, а также с Календарным планом прохождения соответствующего курса - с тем, чтобы иметь возможность вспомнить уже пройденный материал данного курса и на этой основе подготовиться к восприятию новой информации, следуя логике изложения курса преподавателем-лектором.

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

В процессе лекционного занятия студент ведет свой конспект лекций, делая записи, касающиеся основных тезисов лектора. Это могут быть исходные проблемы и вопросы, ключевые понятия и их определения, важнейшие положения и выводы, существенные оценки и т.д.

В заключительной части лекции студент может задать вопросы преподавателю по содержанию лекции, уточняя и уясняя для себя теоретические моменты, которые остались ему непонятными.

Стоит отметить, что необходимо также систематическая самостоятельная работа студента. Самостоятельная работа студента, прежде всего, подразумевает изучение им

учебной и научной литературы, рекомендуемой рабочей программой дисциплины и программой курса. Кроме того, студент должен уделять время детальному и вдумчивому изучению нормативно-правовых документов, регулирующих перемещение товаров через таможенную границу ЕАЭС.

6. Учебная литература и ресурсы информационно-телекоммуникационной сети "Интернет", включая перечень учебно-методического обеспечения для самостоятельной работы обучающихся по дисциплине (модулю)

6.1. Основная литература:

1. Терминосистема таможенного администрирования: методология формирования и практика применения : монография / под общ. ред. д-ра экон. наук, доцента М. В. Бойковой. – Москва : РИО Российской таможенной академии, 2023. - 261 с. – ISBN 978-5-9590-1296-0. - Текст : электронный. - URL: <https://znanium.ru/catalog/product/2202979>
2. Статистический анализ с применением программных средств : учебное пособие / Н. В. Ширкунов, М. М. Цвиль, Е. В. Родительская, И. М. Турланова. – Москва : РИО Российской таможенной академии, 2021. - 141 с. – ISBN 978-5-9590-1222-9. - Текст : электронный. - URL: <https://znanium.ru/catalog/product/2203133> (дата обращения: 28.08.2026).
3. Федоренко К. П. Основы таможенного администрирования: Учебное пособие для вузов / К.П. Федоренко, Ю.А. Борисова, В.А. Фильчакова, Ю.В. Лузгина, А.А. Арошидзе. - Санкт-Петербург : Троицкий мост, 2026. - 212 с. - ISBN 978-5-6055519-0-4. - URL: <https://ibooks.ru/bookshelf/411153/reading> (дата обращения: 28.08.2026). - Текст: электронный.

6.2. Дополнительная литература:

1. Искусственный интеллект: от фундаментальных проблем к прикладным задачам Том 1 : монография / под ред. д.э.н., проф. Е. Н. Макаренко. - Ростов-на-Дону : Издательско-полиграфический комплекс Ростовского государственного экономического университета (РИНХ), 2025. - 394 с. - ISBN 978-5-7972-3390-9. - Текст : электронный. - URL: <https://znanium.ru/catalog/product/2227800>
2. Барило Н. В. Применение искусственного интеллекта в таможенном деле // Океанский менеджмент. 2023. №3 (21). URL: <https://cyberleninka.ru/article/n/primeneniye-iskusstvennogo-intellekta-v-tamozhennom-dele>.
3. Шабля В. О., Коноваленко С. А., Орлов Е. О., Галямин Н. А. МЕТОДЫ СЕМАНТИЧЕСКОГО АНАЛИЗА НА ОСНОВЕ МОДЕЛЕЙ МАШИННОГО ОБУЧЕНИЯ С ИСПОЛЬЗОВАНИЕМ ИСКУССТВЕННЫХ НЕЙРОННЫХ СЕТЕЙ // НИР/S&R. 2025. №1 (21). URL: <https://cyberleninka.ru/article/n/metody-semanticheskogo-analiza-na-osnove-modeley-mashinnogo-obucheniya-s-ispolzovaniem-iskusstvennyh-neyronnyh-setey>.
4. Квасов Михаил Николаевич, Митряев Геннадий Андреевич, Прохоров Михаил Александрович РАЗРАБОТКА МАТЕМАТИЧЕСКОЙ МОДЕЛИ СЕМАНТИЧЕСКОГО АНАЛИЗА И СИНТЕЗА ДАННЫХ // Известия ТулГУ. Технические науки. 2023. №11. URL: <https://cyberleninka.ru/article/n/razrabotka-matematicheskoy-modeli-semanticheskogo-analiza-i-sintez-dannyh>.

6.3 Учебно-методическое обеспечение самостоятельной работы.

Положение об организации самостоятельной работы студентов ФГБОУ ВО «Российская академия народного хозяйства и государственной службы при Президенте Российской Федерации»

Тестовые задания

6.4 Нормативные правовые документы.

6.3. Интернет-ресурсы.

Для освоения дисциплины следует пользоваться доступом через сайт научной библиотеки <http://nwapa.spb.ru/> к следующим подписным электронным ресурсам:

Русскоязычные ресурсы:

- официальный сайт Евразийского экономического союза <http://www.eaeunion.org/>;
- официальный сайт Евразийской экономической Комиссии <http://www.eurasiancommission.org/>;
- электронные учебники электронно-библиотечной системы (ЭБС) «Айбукс»;
- электронные учебники электронно-библиотечной системы (ЭБС) «Лань»;
- статьи из периодических изданий по общественным и гуманитарным наукам «Ист-Вью»
- энциклопедии, словари, справочники «Рубрикон»;
- полные тексты диссертаций и авторефератов Электронная Библиотека Диссертаций РГБ.

Англоязычные ресурсы:

- EBSCO Publishing - доступ к мультидисциплинарным полнотекстовым базам данных различных мировых издательств по бизнесу, экономике, финансам, бухгалтерскому учету, гуманитарным и естественным областям знаний, рефератам и полным текстам публикаций из научных и научно-популярных журналов.

Кроме вышеперечисленных ресурсов, используются следующие ресурсы сети Интернет: <http://uristy.ucoz.ru/>; <http://www.garant.ru/>; <http://www.kodeks.ru/>

6.4. Иные ресурсы.

В ходе образовательного процесса не используется.

7. Материально-техническая база, информационные технологии, программное обеспечение и информационные справочные системы

Информационные средства обучения:

- Системы, используемые для поиска источников информации в сети Интернет;
- Программные задачи компании ООО «СТМ», являющегося разработчиком программного обеспечения для железнодорожной логистики и внешнеэкономической деятельности: «ВЭД-Декларант» - программа для специалистов по таможенному оформлению, непосредственно занимающихся заполнением деклараций на товары, «ВЭД-Инфо» - уникальный электронный справочник, содержащий всю актуальную нормативно-правовую базу в сфере внешнеэкономической деятельности;

Системы дистанционного обучения LMS "Moodlle".

Описание материально-технической базы, необходимой для осуществления образовательного процесса по дисциплине

№ п/п	Наименование
1.	Специализированная аудитория «Информационные технологии в таможенном деле» и компьютерный класс - оснащены 52-мя рабочими станциями ПК, на которых установлены программные средства ВЭД-Декларант, ВЭД-Инфо,

	Тренажер «Прием и регистрация декларации на товары», а также оснащены средствами мультимедиа и 4-мя досками (по 2 в каждом из классов)
2.	Специализированная аудитория «Лаборатория товароведения и экспертизы в таможенном деле» - оснащена средствами мультимедиа, 2-мя досками, демонстрационными материалами, отражающими процессы осуществления таможенного контроля и таможенных операций.