

**Федеральное государственное образовательное
бюджетное учреждение высшего образования**

**«ФИНАНСОВЫЙ УНИВЕРСИТЕТ ПРИ
ПРАВИТЕЛЬСТВЕ РОССИЙСКОЙ ФЕДЕРАЦИИ»
(Финансовый университет)**

**Департамент анализа данных и машинного обучения
Факультета информационных технологий и анализа больших данных**

СОГЛАСОВАНО

АО «ЦПР»

(наименование организации)

Генеральный директор

(должность представителя работодателя)

Самохин А.Н.

(подпись)

18.04.2023 г.

(ФИО)

УТВЕРЖДАЮ

Проректор по учебной и
методической работе

Е.А. Каменева

25.04.2023 г.

С.В. Макрушин

Семантические технологии

Рабочая программа дисциплины

для студентов, обучающихся по направлению подготовки

01.03.02 - Прикладная математика и информатика,

ОП «Прикладная математика и информатика»

(Анализ данных и принятие решений в экономике и финансах),

ОП «Прикладное машинное обучение» (Прикладное машинное обучение)

Рекомендовано Ученым советом

*Факультета информационных технологий и анализа больших данных
(протокол №31 от 18.04.2023г.)*

Одобрено Советом учебно-научного

*Департамента анализа данных и машинного обучения
(протокол №2 от 29.03.2023г.)*

Москва 2023

Содержание

1. Наименование дисциплины.....	2
2. Перечень планируемых результатов освоения образовательной программы (перечень компетенций) с указанием индикаторов их достижения и планируемых результатов обучения по дисциплине	2
3. Место дисциплины в структуре образовательных программ.....	3
4. Объем дисциплины (модуля) в зачетных единицах и в академических часах с выделением объема аудиторной (лекции, семинары) и самостоятельной работы обучающихся	4
5. Содержание дисциплины, структурированное по темам (разделам) дисциплины с указанием их объемов (в академических часах) и видов учебных занятий.....	4
5.1. Содержание дисциплины.....	4
5.2. Учебно-тематический план.....	7
5.3. Содержание семинаров, практических занятий.....	9
6. Перечень учебно-методического обеспечения для самостоятельной работы обучающихся по дисциплине.....	10
6.1. Перечень вопросов, отводимых на самостоятельное освоение дисциплины, формы внеаудиторной самостоятельной работы.....	10
6.2. Перечень вопросов, заданий, тем для подготовки к текущему контролю.....	11
7. Фонд оценочных средств для проведения промежуточной аттестации обучающихся по дисциплине.....	13
8. Перечень основной и дополнительной учебной литературы, необходимой для освоения дисциплины	17
9. Перечень ресурсов информационно-телекоммуникационной сети «Интернет», необходимых для освоения дисциплины.....	18
10. Методические указания для обучающихся по освоению дисциплины.	19
11. Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине, включая перечень необходимого программного обеспечения и информационных справочных систем.....	20
12. Описание материально-технической базы, необходимой для осуществления образовательного процесса по дисциплине.....	21

1. Наименование дисциплины

«Семантические технологии».

2. Перечень планируемых результатов освоения образовательной программы (перечень компетенций) с указанием индикаторов их достижения и планируемых результатов обучения по дисциплине

Код компетенции	Наименование компетенции	Индикаторы достижения компетенции	Результаты обучения (умения и знания), соотнесенные с индикаторами достижения компетенции
ОП «Прикладная математика и информатика» (Анализ данных и принятие решений в экономике и финансах)			
ОПК-4	Способен понимать принципы работы современных информационных технологий и использовать их для решения задач профессиональной деятельности	1.Использует информационно-коммуникационные технологии при поиске необходимой информации, сбора, визуализации и обработки данных.	Знать: принципы семантических технологий. Уметь: использовать семантические технологии.
		2.Осуществляет рациональный выбор программного продукта в зависимости от поставленной задачи	Знать: основные программные решения для использования семантических технологий. Уметь: выполнять рациональный выбор программных решений в области семантических технологий.
		3. Владеет навыками обеспечения информационной безопасности автоматизированных систем.	Знать: способы обеспечения информационной безопасности. Уметь: использовать средства обеспечения информационной безопасности.
ПКП-5	Способность применять методы и инструменты анализа данных и машинного обучения при подготовке аналитического обоснования финансово-экономических	1. Осуществляет обоснованный выбор методов и инструментов обработки данных для решения экономических и финансовых задач.	Знать: основные программные и алгоритмические решения для использования семантических технологий в приложениях. Уметь: выполнять рациональный выбор алгоритмических и программных решений в области семантических технологий в приложениях.

	решений	2. Владеет навыками решения прикладных задач с использованием методов и инструментов анализа данных и машинного обучения.	<u>Знать:</u> способы решения прикладных задач с использованием семантических технологий. <u>Уметь:</u> использовать семантические технологии для решения прикладных задач.
ОП «Прикладное машинное обучение» (Прикладное машинное обучение)			
ПКП-3	Способность выполнять структурирование данных, выявлять признаки ценные для моделей машинного обучения.	1. Демонстрирует знание принципов процесса структурирования данных и формирования признаков для моделей машинного обучения.	<u>Знать:</u> возможности семантических технологий для структурированной данных. <u>Уметь:</u> использовать семантические технологии для структурированной данных.
		2. Применяет инструментальные средства для преобразования данных и структурирования данных в рамках их предобработки для использования в моделях машинного обучения.	<u>Знать:</u> инструментальные средства для структурирования данных с помощью семантических технологий. <u>Уметь:</u> использовать инструментальные средства для структурирования данных с помощью семантических технологий.
		3. Владеет практическим навыком инжиниринга признаков для моделей машинного обучения с учетом структуры исходных данных и предметной области прикладной задачи.	<u>Знать:</u> подходы для построения эмбединга графов знаний. <u>Уметь:</u> строить эмбединги графов знаний.

3. Место дисциплины в структуре образовательных программ

Дисциплина «Семантические технологии» относится к Циклу профиля (элективный) по направлению подготовки 01.03.02 – Прикладная математика и информатика, ОП «Прикладная математика и информатика» (Анализ данных и принятие решений в экономике и финансах).

Дисциплина «Семантические технологии» является дисциплиной Профиля Прикладное машинное обучение по направлению подготовки 01.03.02 – Прикладная математика и информатика, ОП «Прикладное машинное обучение».

4. Объем дисциплины(модуля) в зачетных единицах и в академических часах с выделением объема аудиторной (лекции, семинары) и самостоятельной работы обучающихся

ОП «Прикладная математика и информатика» (Анализ данных и принятие решений в экономике и финансах)

Вид учебной работы по дисциплине	Всего (в з/е и часах)	Семестр 7 (в часах)
Общая трудоемкость дисциплины	3/108	108
<i>Контактная работа – Аудиторные занятия</i>	50	50
<i>Лекции</i>	16	16
<i>Семинары, практические занятия</i>	34	34
Самостоятельная работа	58	58
Вид текущего контроля	Контрольная работа	Контрольная работа
Вид промежуточной аттестации	Экзамен	Экзамен

ОП «Прикладное машинное обучение» (Прикладное машинное обучение)

Вид учебной работы по дисциплине	Всего (в з/е и часах)	Семестр 7 (в часах)
Общая трудоемкость дисциплины	4/144	144
<i>Контактная работа – Аудиторные занятия</i>	50	50
<i>Лекции</i>	16	16
<i>Семинары, практические занятия</i>	34	34
Самостоятельная работа	94	94
Вид текущего контроля	Контрольная работа	Контрольная работа
Вид промежуточной аттестации	Экзамен	Экзамен

5. Содержание дисциплины, структурированное по темам (разделам) дисциплины с указанием их объемов (в академических часах) и видов учебных занятий

5.1. Содержание дисциплины

Тема 1. Графы знаний, семантические технологии

Графы знаний. Модели данных, используемые для представления информации и структурированных знаний. Представление сложных взаимосвязей между данными, связей сущностей, атрибутов и отношений. Семантические сети. Формализация с помощью графов знаний. Понятие онтологии при создании графов

знаний. Использование графов знаний и онтологий в обработке текстов на естественном языке и приложениях искусственного интеллекта.

Технологии Semantic Web. Введение в Semantic Web, технологический стек Semantic Web. Структурирование данных с помощью RDF, выполнение запросов к Semantic Web с помощью SPARQL, использование языков запросов и правил. Обзор архитектуры Semantic Web, интеграции RDF и систем, основанных на правилах, обзор языков онтологий и управление онтологиями.

Тема 2. Онтологии, связанные данные

Понятие онтологии при создании графов знаний. Использование графов знаний и онтологий в обработке текстов на естественном языке и приложениях искусственного интеллекта. Построение онтологий для Semantic Web, язык OWL. Поддержка алгоритмов формальных рассуждений в Semantic Web. Использование публичных онтологий, таких как Shema.org, FOAF и других. Принципы и инструменты конструирования собственных онтологий.

Linked Data как способ соединения различных источников данных из разных местоположений с помощью технологий Semantic Web.

Технология связанных данных (linked data) как подход, основывающийся на стандартных веб-технологиях, предназначенный для реализации технологического стека «Семантического Веба». Цель связанных данных – представление данных для решения задач автоматизации обработки данных и описания структуры данных нежеле просто размещение данных в вебе. Обязательными составляющими технологии связанных данных являются технологии адресации и использование схемы данных. Технология связанных данных помогает предоставлять данные в виде графа, который может быть легко объединен с другими источниками данных. Открытые описания схем данных на основе технологий Semantic Web, знакомство с Schema.org. Построение общих схем для данных и их использование для объединения различных источников данных на основе технологий Linked Data.

Тема 3. Крупные универсальные базы знаний

Крупномасштабные универсальные базы знаний, кейсы Wikidata и DBPedia. Сравнение Wikidata, DBPedia и Википедии. Использование Wikidata как источника данных.

Язык запросов SPARQL, использование открытых эндпоинтов для получения информации из крупных универсальных баз знаний. Онтология Wikidata.

Тема 4. Базы знаний

Базы знаний как хранилища структурированных данных, позволяющие эффективно извлекать и анализировать факты по данному предмету. Машино-читаемые базы знаний. Базы знаний, построенные на технологиях Semantic Web. Сравнительный анализ существующих решений. Организация машино-читаемых баз знаний. Построение запросов к базе знаний с использованием языка запросов для Semantic Web SPARQL. Построение корпоративных и отраслевых баз знаний.

Тема 5. Векторные представления графов знаний

Принципы построения эмбедингов. Эмбединги для графов, node2vec. Эмбединги для графов знаний. Типы графовых эмбедингов: билинейные (DistMult, ComplEx), не билинейные (HolE), геометрические: трансляционные эмбединги (TransE, TransH), трансляционные с дополнительными эмбедингами, ротационные (RotatE), глубокие эмбединги (ConvE). Примеры использования эмбедингов.

Тема 6. Правила вывода и обогащение графов знаний

Вывод на основе правил и на основе онтологий. Принципы вывода на основе правил, языки описания правил. Правила вывода в языках OWL, RDF Schema. Правила вывода в SPARQL. Создание новых фактов на основе вывода. Возможность создания новых фатков на основе эмбедингов.

Тема 7. Языковые модели и семантика в естественном языке

Задача построения языковых моделей. Распределение n-грамм в естественном языке. Марковские процессы. Дистрибутивная гипотеза. Матрица совместной встречаемости, модель TF-IDF. Тематическое моделирование. Модели word2vec и BERT. Модели GPT. Семантика понятий в естественном языке и языковые модели.

Тема 8. Автоматическое извлечение семантики из текста на естественном языке

Построение графов знаний и аналогичных конструкций на основе автоматического анализа текста. Задача выявления именованных сущностей в тексте. Задача выявления отношений в тексте. Построение нестрогих семантических графов.

5.2. Учебно-тематический план

ОП «Прикладная математика и информатика» (Анализ данных и принятие решений в экономике и финансах)

№ п/п	Наименование тем (разделов) дисциплины	Трудоемкость в часах					Формы текущего контроля успеваемости
		Всего	Контактная работа - Аудиторная работа			Самостоя тельная работа	
			Общая, в т.ч.:	Лекции	Семинары, практические занятия		
1.	Графы знаний, семантические технологии	12	6	2	4	6	Самостоятельн ые работы. Участие в решении задач на практических занятиях. Собеседования по домашним заданиям.
2.	Онтологии, связанные данные	14	6	2	4	8	
3.	Крупные универсальные базы знаний	14	6	2	4	8	
4.	Базы знаний	14	6	2	4	8	
5.	Векторные представления графов знаний	14	6	2	4	8	
6.	Правила вывода и обогащение графов знаний	14	6	2	4	8	
7.	Языковые модели и семантика в естественном языке	14	8	2	6	6	
8.	Автоматическое извлечение семантики из текста на естественном языке	12	6	2	4	6	
	В целом по дисциплине	108	50	16	34	58	Согласно учебному плану: контрольная работа
	Итого в %		46	32	68	54	

ОП «Прикладное машинное обучение» (Прикладное машинное обучение)

№ п/п	Наименование тем (разделов) дисциплины	Трудоемкость в часах					Формы текущего контроля успеваемости
		Всего	Контактная работа - Аудиторная работа			Самостоя тельная работа	
			Общая, в т.ч.:	Лекции	Семинары, практические занятия		
1.	Графы знаний, семантические технологии	16	6	2	4	10	Самостоятельн ые работы. Участие в решении задач на практических занятиях. Собеседования по домашним заданиям.
2.	Онтологии, связанные данные	18	6	2	4	12	
3.	Крупные универсальные базы знаний	18	6	2	4	12	
4.	Базы знаний	18	6	2	4	12	
5.	Векторные представления графов знаний	18	6	2	4	12	
6.	Правила вывода и обогащение графов знаний	18	6	2	4	12	
7.	Языковые модели и семантика в естественном языке	20	8	2	6	12	
8.	Автоматическое извлечение семантики из текста на естественном языке.	18	6	2	4	12	
	В целом по дисциплине	144	50	16	34	94	Согласно учебному плану: контрольная работа
	Итого в %		35	32	68	65	

5.3. Содержание семинаров, практических занятий

Наименование тем (разделов) дисциплины	Перечень вопросов для обсуждения на семинарских, практических занятиях, рекомендуемые источники из разделов 8,9 (указывается раздел и порядковый номер источника)	Формы проведения занятий
Графы знаний, семантические технологии	Построение семантической сети. Запись RDF в формате Turtle. Разворачивание и базовое использование RDF-хранилища. <i>Рекомендуемые источники: п.8, [1]-[5]</i>	Интерактивная форма, Практикум по решению задач по тематике занятия в малых группах (2-4 студента) и коллективное обсуждение решений
Онтологии, связанные данные	Знакомство с публичными онтологиями. Не формализованное описание онтологии. Запись онтологии в формате RDF Schema. <i>Рекомендуемые источники: п.8, [1]-[5]</i>	Интерактивная форма, Практикум по решению задач по тематике занятия в малых группах (2-4 студента) и коллективное обсуждение решений
Крупные универсальные базы знаний	Изучение функциональных возможностей и структуры Wikidata. Базовое использование эндпоинта Wikidata. Написание запросов на SPARQL к Wikidata. <i>Рекомендуемые источники: п.8, [1]-[5]</i>	Интерактивная форма, Практикум по решению задач по тематике занятия в малых группах (2-4 студента) и коллективное обсуждение решений
Базы знаний	Формирование схемы и наполнения базы знаний по определенному домену. <i>Рекомендуемые источники: п.8, [1]-[5]</i>	Интерактивная форма, Практикум по решению задач по тематике занятия в малых группах (2-4 студента) и коллективное обсуждение решений
Векторные представления графов знаний	Построение векторного представления терминов графов знаний при помощи трансляционных моделей (TransE). <i>Рекомендуемые источники: п.8, [1]-[5]</i>	Интерактивная форма, Практикум по решению задач по тематике занятия в малых группах (2-4 студента) и коллективное обсуждение решений

Правила вывода и обогащение графов знаний	Описание и применение вывода на основе правил. <i>Рекомендуемые источники: п.8, [1]-[5]</i>	Интерактивная форма, Практикум по решению задач по тематике занятия в малых группах (2-4 студента) и коллективное обсуждение решений
Языковые модели и семантика в естественном языке	Определение семантического отношения слов с помощью предобученных языковых моделей (word2vec, BERT). <i>Рекомендуемые источники: п.8, [1]-[5]</i>	Интерактивная форма, Практикум по решению задач по тематике занятия в малых группах (2-4 студента) и коллективное обсуждение решений
Автоматическое извлечение семантики из текста на естественном языке.	Решение задачи выявления именованных сущностей из текста. Решение задачи выявления отношения между сущностями из текста. <i>Рекомендуемые источники: п.8, [1]-[5]</i>	Интерактивная форма, Практикум по решению задач по тематике занятия в малых группах (2-4 студента) и коллективное обсуждение решений

6. Учебно-методическое обеспечение для самостоятельной работы обучающихся по дисциплине

6.1. Перечень вопросов, отводимых на самостоятельное освоение дисциплины, формы внеаудиторной самостоятельной работы

Наименование тем (разделов) дисциплины	Перечень вопросов, отводимых на самостоятельное освоение	Формы внеаудиторной самостоятельной работы
Графы знаний, семантические технологии	Загрузка графа знаний в формате RDF Turtle в хранилище.	Работа с учебной литературой. Решение типовых задач. Разбор вопросов по теме занятия. Выполнение домашних заданий к каждому занятию.
Онтологии, связанные данные	Запись онтологии в формате OWL. Применение онтологии в RDF-хранилищ	Работа с учебной литературой. Решение типовых задач. Разбор вопросов по теме занятия. Выполнение домашних заданий к каждому занятию.

Крупные универсальные базы знаний	Автоматическая выгрузка данных из Wikidata через API с помощью SPARQL запросов.	Работа с учебной литературой. Решение типовых задач. Разбор вопросов по теме занятия. Выполнение домашних заданий к каждому занятию.
Базы знаний	Формирование программной обвязки, использующей данные из базы знаний через API и SPARQL запросы.	Работа с учебной литературой. Решение типовых задач. Разбор вопросов по теме занятия. Выполнение домашних заданий к каждому занятию.
Векторные представления графов знаний	Построение векторного представления терминов графов знаний при помощи иных типов моделей (RotatE, DistMult).	Работа с учебной литературой. Решение типовых задач. Разбор вопросов по теме занятия. Выполнение домашних заданий к каждому занятию.
Правила вывода и обогащение графов знаний	Описание и применение вывода на основе онтологий (таких как OWL).	Работа с учебной литературой. Решение типовых задач. Разбор вопросов по теме занятия. Выполнение домашних заданий к каждому занятию.
Языковые модели и семантика в естественном языке	Определение семантического отношения слов с помощью больших предобученных языковых моделей (T5).	Работа с учебной литературой. Решение типовых задач. Разбор вопросов по теме занятия. Выполнение домашних заданий к каждому занятию.
Автоматическое извлечение семантики из текста на естественном языке.	Решение задачи построения нестроого графа знаний на основе текста.	Интерактивная форма, Практикум по решению задач по тематике занятия в малых группах (2-4 студента) и коллективное обсуждение решений

6.2. Перечень вопросов, заданий, тем для подготовки к текущему контролю

Примерный перечень вопросов к контрольной работе

1. В чём состоит отличие графов знаний от других видов графов?
2. Какие описываются правила вывода для графов знаний?
3. Как строится функция ошибки при построении эмбединга TransE?
4. Как строится функция ошибки при построении эмбединга TransH?
5. Как строится эмбединг DistMult?
6. Для каких задач может использоваться векторное представление графов знаний?

7. Какие задачи можно решать с помощью языковой модели?
8. Как могут автоматически выявляться семантические отношения между словами?
9. Для чего могут использоваться открытые онтологии?
10. В чем специфика крупномасштабных универсальных баз знаний?
11. На каких задачах выполняют обучение векторных представлений графов знаний?
12. Какие задачи решает языковая модель?
13. Какие подходы используются для автоматического извлечения семантики слов из текста на естественном языке?
14. Что такое «связанные данные» и какой прикладной смысл их создания?
15. Какие приложения можно построить на основе базы знаний?

Примерные задания контрольной работы

1. Опишите с помощью RDF Turtle базу знаний об учебном процессе в университете.
2. С помощью RDF Schema постройте онтологию для библиотеки.
3. Напишите SPARQL запрос для Wikidata для извлечения информации о всех фильмах Кристофера Нолана.
4. Опишите с помощью RDF Turtle базу знаний о школе.
4. Напишите SPARQL запрос для Wikidata для извлечения информации о всех моделях автомобилей созданных на Волжском автозаводе.
5. Постройте онтологию, на основе которой можно построить вывод, опишите принцип работы вывода.
6. Решите задачу извлечения именованных сущностей из текста на естественном языке

Критерии балльной оценки различных форм текущего контроля успеваемости содержатся в соответствующих методических рекомендациях Департамента анализа данных и машинного обучения Факультета информационных технологий и анализа больших данных.

7. Фонд оценочных средств для проведения промежуточной аттестации обучающихся по дисциплине

Перечень компетенций с указанием индикаторов их достижения в процессе освоения образовательной программы содержится в разделе 2. «Перечень планируемых результатов освоения образовательной программы (перечень компетенций) с указанием индикаторов их достижения и планируемых результатов обучения по дисциплине».

Типовые контрольные задания или иные материалы, необходимые для оценки индикаторов достижения компетенций, умений и знаний

Наименование компетенции	Наименование индикаторов достижения компетенции	Результаты обучения (умения и знания), соотнесенные с индикаторами достижения компетенции	Типовые контрольные задания
ОП «Прикладная математика и информатика» (Анализ данных и принятие решений в экономике и финансах)			
ОПК-4 Способен понимать принципы работы современных информационных технологий и использовать их для решения задач профессиональной деятельности	1.Использует информационно-коммуникационные технологии при поиске необходимой информации, сбора, визуализации и обработки данных.	Знать: принципы семантических технологий. Уметь: использовать семантические технологии.	Создайте небольшую (до 100 триплетов) базу знаний для ресторанного бизнеса.
	2.Осуществляет рациональный выбор программного продукта в зависимости от поставленной задачи	Знать: основные программные решения для использования семантических технологий. Уметь: выполнять рациональный выбор программных решений в области семантических технологий.	Выберите RDF хранилище для создания корпоративной базы знаний средних размеров, поддерживающей правила вывода и RDF.
	3.Владеет навыками обеспечения информационной безопасности автоматизированных систем.	Знать: способы обеспечения информационной безопасности. Уметь: использовать средства обеспечения информационной безопасности.	Как защитить базу знаний от несанкционированного доступа?

ПКП-5 Способность применять методы и инструменты анализа данных и машинного обучения при подготовке аналитического обоснования финансово-экономических решений	1.Осуществляет обоснованный выбор методов и инструментов обработки данных для решения экономических и финансовых задач.	Знать: основные программные и алгоритмические решения для использования семантических технологий в приложениях. Уметь: выполнять рациональный выбор алгоритмических и программных решений в области семантических технологий в приложениях.	Предложите процедуру выбора эмбединга графа знаний для решения прикладной задачи поиска близких сущностей из базы знаний.
	2.Владеет навыками решения прикладных задач с использованием методов и инструментов анализа данных и машинного обучения.	Знать: способы решения прикладных задач с использованием семантических технологий. Уметь: использовать семантические технологии для решения прикладных задач.	Постройте онтологию базы знаний для компании продающей ноутбуки.
ОП «Прикладное машинное обучение» (Прикладное машинное обучение)			
ПКП-3 Способность выполнять структурирование данных, выявлять признаки ценные для моделей машинного обучения.	1.Демонстрирует знание принципов процесса структурирования данных и формирования признаков для моделей машинного обучения.	Знать: возможности семантических технологий для структурированной данных. Уметь: использовать семантические технологии для структурированной данных.	Постройте онтологию для предметной области «Кинофильмы», позволяющую структурировать данные о подборке фильмов онлайн-кинотеатра.
	2.Применяет инструментальные средства для преобразования данных и структурирования данных в рамках их предобработки для использования в моделях машинного обучения.	Знать: инструментальные средства для структурирования данных с помощью семантических технологий. Уметь: использовать инструментальные средства для структурирования данных с помощью семантических технологий.	Решите задачу поиска похожих сущностей базы знаний с помощью трансляционного эмбединга базы знаний.
	3.Владеет практическим навыком инжиниринга	Знать: подходы для построения эмбединга графов знаний.	Постройте эмбединги для базы знаний среднего размера

	признаков для моделей машинного обучения с учетом структуры исходных данных и предметной области прикладной задачи.	Уметь: строить эмбединги графов знаний.	(порядка 10 тысяч триплетов).
--	---	--	-------------------------------

Примерные вопросы для подготовки к экзамену

1. Семантические сети. Графы знаний. Представление сложных взаимосвязей между данными, связей сущностей, атрибутов и отношений.
2. Основные технологии Semantic Web.
3. RDF. Формат RDF Turtle. Пример описания базы знаний с помощью RDF Turtle.
4. Понятие онтологии при создании графов знаний. Ключевые технологии для описания онтологий в Semantic Web.
5. RDF Schema. Сравнение с OWL. Пример описания онтологии предметной области с помощью RDF Schema.
6. OWL. Сравнение с RDF Schema. Пример описания онтологии предметной области с помощью OWL.
7. Открытые онтологии. Примеры открытых онтологий. Принципы построения частной онтологии с использованием открытой онтологии.
8. Крупномасштабные универсальные базы знаний, кейсы Wikidata и DBPedia. Сравнение Wikidata, DBPedia и Википедии.
9. Крупномасштабная универсальная база знаний Wikidata. Общее описание, структура, функциональные возможности. Связь между Wikidata и Wikipedia.
10. Язык запросов SPARQL, использование открытых эндпоинтов для получения информации из крупных универсальных баз знаний с помощью запросов на SPARQL.
11. Базы знаний – определения и альтернативные принципы построения. Машиночитаемые базы знаний. Возможные инструменты для построения машиночитаемых баз знаний.

12. Принципы построения эмбедингов и построение эмбедингов для графов.
Эмбединг node2vec.
13. Эмбединги для графов знаний. Общая классификация и особенности.
14. Билинейные эмбединги для графов знаний DistMult, ComplEx.
15. Трансляционные эмбединги для графов знаний TransE, TransH.
16. Глубокие эмбединги для графов знаний ConvE.
17. Вывод в графах знаний. Вывод на основе правил.
18. Вывод в графах знаний. Вывод на основе онтологии.
19. Языковые модели. Определение и цели языковых моделей, основные подходы к построению.
20. Дистрибутивная гипотеза, построения эмбедингов слов с помощью word2vec.
Семантический анализ с помощью word2vec.
21. Современные глубокие модели построения языковых моделей на основе трансформеров.
22. Автоматический анализ текста на естественном языке для обогащения графов знаний. Выявление именованных сущностей.
23. Автоматический анализ текста на естественном языке для обогащения графов знаний. Выявление именованных сущностей.
24. Автоматический анализ текста на естественном языке для обогащения графов знаний. Выявление отношений.
25. Обогащение графов знаний на основе автоматического анализа текстов на естественном языке. Построение нестрогих графов знаний.

Пример экзаменационного билета

Федеральное государственное образовательное бюджетное учреждение
высшего образования

«ФИНАНСОВЫЙ УНИВЕРСИТЕТ ПРИ ПРАВИТЕЛЬСТВЕ
РОССИЙСКОЙ ФЕДЕРАЦИИ»
(Финансовый университет)

Департамент анализа данных и машинного обучения

Дисциплина **Семантические технологии**

Факультет информационных технологий и анализа больших данных

Форма обучения **очная**

Семестр: **7 семестр**

Направление подготовки: **01.03.02 - Прикладная математика и информатика**

Профиль:

ЭКЗАМЕНАЦИОННЫЙ БИЛЕТ №

1 вопрос (**20 баллов**) RDF. Формат RDF Turtle. Пример описания базы знаний с помощью RDF Turtle.

2 вопрос (**20 баллов**) С помощью RDF Schema постройте онтологию для библиотеки.

3 вопрос (**20 баллов**) Для графа знаний из файла kg1.zip построить эмбединги сущностей с помощью метода TransE.

Подготовил: _____ Макрушин С.В.

Утверждаю:

Первый заместитель

руководителя департамента

_____ ФИ.О.

Дата _____

8. Перечень основной и дополнительной учебной литературы, необходимой для освоения дисциплины

Основная литература:

[1] Криволапов, С. Я. Математика на Python : учебник / С. Я. Криволапов, М. Б. Хрипунова. — Москва : КноРус, 2022. — 455. — ЭБС BOOK.ru. - URL: <https://book.ru/book/943665> (дата обращения: 15.06.2023). — Текст : электронный.

[2] Криволапов, С. Я. Использование языка Python в теории вероятностей : учебник / С. Я. Криволапов. — Москва : Прометей, 2022. — 492 с. — ЭБС Лань. — URL: <https://e.lanbook.com/book/220814> (дата обращения: 15.06.2023). — Текст : электронный.

Дополнительная литература:

[3] Криволапов, С. Я. Статистические вычисления на платформе Jupyter Notebook с использованием Python: учебник / С. Я. Криволапов. — Москва : КноРус, 2022. — 431 с. — ЭБС BOOK.RU. - URL: <https://book.ru/book/943660> (дата обращения: 15.06.2023). — Текст : электронный.

[4] Рутковская, Д. Нейронные сети, генетические алгоритмы и нечеткие системы / Д. Рутковская, М. Пилиньский, Л. Рутковский; пер. с польск. И. Д. Рудинского. - 2-е изд., стереотип. — Москва : Гор. линия-Телеком, 2013. - 384 с. - ЭБС ZNANIUM.com. - URL: <https://new.znanium.com/catalog/product/414545> (дата обращения: 15.06.2023). - Текст : электронный.

[5] Ростовцев, В. С. Искусственные нейронные сети : учебник / В. С. Ростовцев. — Санкт-Петербург : Лань, 2021. — 216 с. — ЭБС Лань. — URL: <https://e.lanbook.com/reader/book/160142> (дата обращения: 15.06.2023). — Текст : электронный.

9. Перечень ресурсов информационно-телекоммуникационной сети «Интернет», необходимых для освоения дисциплины

1. Электронная библиотека Финансового университета (ЭБ) <http://elib.fa.ru/>
2. Электронно-библиотечная система BOOK.RU <http://www.book.ru>
3. Электронно-библиотечная система «Университетская библиотека ОНЛАЙН» <http://biblioclub.ru/>
4. Электронно-библиотечная система Znanium <http://www.znanium.com>
5. Электронно-библиотечная система издательства «ЮРАЙТ» <https://urait.ru/>
6. Электронно-библиотечная система издательства Проспект <http://ebs.prospekt.org/books>

7. Электронно-библиотечная система издательства Лань <https://e.lanbook.com/>
8. Деловая онлайн-библиотека Alpina Digital <http://lib.alpinadigital.ru/>
9. Электронная библиотека Издательского дома «Гребенников»
<https://grebennikon.ru/>
10. Научная электронная библиотека eLibrary.ru <http://elibrary.ru>
11. Национальная электронная библиотека <http://нэб.рф/>
12. Финансовая справочная система «Финансовый директор» <http://www.1fd.ru/>
13. Электронная документация по работе в pytorch (вариант 1):
<https://neurohive.io/ru/tutorial/glubokoeobuchenie-s-pytorch/>
14. Электронная документация по работе в pytorch (вариант 2):
<https://pytorch.org/tutorials/>
15. Дистрибутив anaconda: <http://anaconda.org/>
16. Электронная документация по работе в библиотеке pandas:
<http://pandas.pydata.org/>
17. Курс лекций по машинному и глубокому обучению (электронный учебник):
<http://www.machinelearning.ru/wiki/images/6/6d/Voron-ML-1.pdf>

10. Методические указания для обучающихся по освоению дисциплины.

Основные этапы работы студента по дисциплине *Семантические технологии*:

1. Предварительная ориентировка в подлежащем изучению учебном материале по программе.
2. Ознакомление с рекомендованной учебной литературой.
3. Слушание и конспектирование лекций, а также выполнение других видов учебной работы.
4. Планирование самостоятельной работы.
5. Обобщение и систематизация информации, почерпнутой из лекций и прочитанной литературы.
6. Выполнение контрольной работы.

Рекомендации по работе с учебным материалом:

1. Осознавайте наличный уровень полученных вами знаний.
2. В ситуации непонимания нужно выявить тот первичный уровень и факторы непонимания, которые стали препятствием понимания последующего.
3. Задавайте сами себе вопросы и пытайтесь ответить на них.

Рекомендации по работе на лекции и с лекционным материалом:

1. Основная задача на лекции – осмысление излагаемого в ней материала. Для этого необходимо слушать лекцию с самого начала, не упуская общих, ориентирующих в материале рассуждений и установок лектора.
2. Ведение записей на лекции важно и полезно для лучшего осмысливания материала, для сохранения информации, с целью ее дальнейшего использования.
3. Для облегчения записи рекомендуется применять сокращения повторяющихся терминов или хорошо известных понятий.

Рекомендации по работе с литературой:

1. Если возникли затруднения при разыскивании материала, по какому-либо конкретному вопросу, следует обратиться к предметному указателю, напечатанному, как правило, в конце каждого литературного источника.
2. Предметный указатель – это алфавитный список основных научных понятий (терминов), содержание которых раскрыто в книге, рядом с термином стоят числа, обозначающие номера страниц, на которых изложен материал, относящийся к данному понятию.

11. Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине, включая перечень необходимого программного обеспечения и информационных справочных систем

11.1. Комплект лицензионного программного обеспечения:

1. Пакет офисных программ
2. Антивирус Kaspersky

11.2. Современные профессиональные базы данных и информационные справочные системы:

1. Информационно-правовая система «Гарант»
2. Информационно-правовая система «Консультант Плюс»
3. Электронная энциклопедия: <http://ru.wikipedia.org/wiki/Wiki>
4. Система комплексного раскрытия информации «СКРИН» - <http://www.skrin.ru/>

11.3. Сертифицированные программные и аппаратные средства защиты информации:

- не предусмотрены.

11.4. Язык программирования Python 3.x в среде Windows.

11.5. Платформа для научных исследований, основанная на языке программирования Python, Anaconda.

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

Наличие аудитории, оснащенной компьютерной техникой и проектором, с возможностью подключения к сети «Интернет».