

Аннотация дисциплины Б.1.2.19 Дисциплина. Проектирование баз знаний

Дисциплина "Проектирование баз знаний" изучается обучающимися по основной профессиональной образовательной программе "Искусственный интеллект в бизнес-аналитике" направления подготовки "09.03.03 Прикладная информатика".

Дисциплина изучается в 8 семестре. Общая трудоемкость дисциплины составляет 108/3 часов/з.ед. Самостоятельная работа заключается в выполнении работ, указанных в разделе 4.

В ходе изучения дисциплины осуществляется текущий контроль в форме технологии рейтингового контроля в соответствии с технологической карты дисциплины, размещенной на электронном курсе, а также промежуточный контроль в форме зачет.

Целью изучения дисциплины является формирование следующих компетенций:

1. FC-2 Способен проводить фронтальные исследования в области фундаментальных и генеративных моделей
2. LLM-3 Проектирует и применяет техники расширения контекста генерации (RAG)
3. О-1 Способен осуществлять управление знаниями, в том числе с применением алгоритмов интеллектуального поиска решений и формирования стратегий
4. ПК-1 Способность проводить обследование организаций, выявлять информационные потребности пользователей, формировать требования к информационной системе
5. ПК-3 Способность проектировать ИС по видам обеспечения

В ходе изучения дисциплины последовательно рассматриваются темы:

1. Лек1 Что такое Knowledge Graph (KG). отличие от традиционных БД. Почему важны KG. Примеры KG
RDF-графы vs Property Graphs. Типовые задачи: хранение знаний, entity linking, reasoning и др.
2. Лек2 Основы моделирования KG: узлы, отношения, свойства, entities, relationships, events, roles.
Пример KG "Кино и актеры"
3. Лек3 Знакомство с Neo4j. Cypher - язык запросов, ч1. MATCH / CREATE / MERGE / DELETE.
WHERE, RETURN, UNWIND, FOREACH.
4. Лек4 Язык запросов Cypher, ч2. Агрегация ('COUNT', 'SUM'), сортировка ('ORDER BY'), пагинация ('SKIP', 'LIMIT').
5. Лек5 Работа с Neo4j из Python: библиотека py2neo
Введение в py2neo: Установка, подключение к базе данных. Основные объекты: 'Graph', 'Node', 'Relationship', 'Subgraph'. Операции: CRUD.
6. Лек6 Написание скриптов на Питоне для операций CRUD. Источники данных: (CSV, JSON, БД) и неструктурированные данные (текст). Извлечение сущностей и связей (NER & RE): обзор методов. ETL-пайплайны для графов.
7. Лек.7 Рекомендательные системы на основе графовых баз знаний.
8. Лек.8 Извлечение новых знаний из графовых баз знаний (пример с медикаментами)
9. Лек.9 Введение в RAG . Преимущества RAG систем.
Архитектура RAG систем. Локальные LLM и Ollama
10. Лек10 Сборка полного RAG-конвейера (без графов). Проектирование пайплайна: Разбиение текста (Text Splitting), создание эмбеддингов, сохранение в Chroma, поиск, генерация.
11. Лекция11 Интеграция RAG и Графов Знаний (Гибридный поиск). Проблемы векторного поиска. Гибридный подход: Совмещение семантического поиска (Chroma) и структурного поиска по графу (Neo4j).
12. Лекция12 Реализация гибридной системы RAG + KG. Итоги курса.

Основными стратегическими образовательными технологиями являются: исследовательские, лекционные занятия, практические и лабораторные занятия. В рамках указанных технологий применяются тактические образовательные технологии: задания, классическая лекция, мини-проекты, семинары с ит-практиками, конкурсы, хакатоны, кейс-чемпионаты, мастер-классы, командные проектные задания.