Assignment 1: QA System and Knowledge Graph Specification

DeFi Protocol Explorer

Evangelos Rafail Kalemkeridis
MSc in AI and Data Science, Athens University of Economics and Business
Instructor: Panos Alexopoulos

2025–2026

Task 1: Select a Scenario and a Domain

Scenario: A GraphRAG-based Question Answering (QA) system designed to allow users to query a knowledge graph of decentralized finance (DeFi) protocols, their blockchains, categories, and relationships such as partnerships, audits, and shared assets.

Domain: The DeFi sector within the cryptocurrency industry. The system focuses on the interconnections among protocols (e.g., Uniswap, Aave, Lido), blockchains (e.g., Ethereum, Arbitrum, Polygon), tokens, and organizations. Data sources include public APIs and repositories such as DeFiLlama, CoinMarketCap, Messari, and The Graph protocol subgraphs.

Task 2: Describe the Envisioned QA System

System name: DeFi Protocol Explorer

The envisioned system provides a natural language interface for querying structured knowledge about DeFi protocols. Users can ask questions about TVL (Total Value Locked), categories, supported tokens, founders, or relationships between projects.

Users:

- Crypto analysts and researchers seeking protocol insights

- Investors and portfolio managers analyzing ecosystems
- Blockchain developers exploring integrations
- Risk and compliance teams assessing exposure

Value: The system aggregates and structures scattered DeFi information, enabling fast, explainable, and context-rich answers. It supports both factual retrieval (e.g., "What is the TVL of Aave?") and analytical reasoning (e.g., "Which blockchains host the most DeFi protocols?").

Task 3: Define the System's Question Space

The DeFi Protocol Explorer supports a wide range of questions from factual to relational and analytical.

Factual Questions

1. What is the TVL of Aave?
2. On which blockchain is Uniswap deployed?
3. Who are the founders of Lido?
4. Which tokens does Curve support?
5. When was Compound launched?

Relational Questions

6. Which protocols are deployed on both Ethereum and Arbitrum?
7. Which DeFi protocols share the same audit firm as SushiSwap?
8. Which projects have partnerships with Chainlink?
9. Which stablecoins are supported by more than five lending platforms?
10. Which protocols belong to the liquid staking category?

Analytical and Summarizing Questions

11. Which blockchain hosts the most DeFi protocols?
12. Which categories have the highest average TVL?
13. How many DeFi protocols were launched after 2021?
14. Which protocols grew their TVL by more than 50% last quarter?
15. What are the top five DeFi protocols by number of users?
16. Which protocols have active governance tokens?
17. Which protocols have bridges to at least three other blockchains?
18. Which DeFi categories are shrinking in TVL?
19. Which protocols are backed by the same venture capital firm?
20. Which projects have migrated from Ethereum to their own chain?

Task 4: Derive Requirements for the Knowledge Graph

To support the question space above, the knowledge graph must represent a set of entities, attributes, and relationships that describe the DeFi ecosystem.

Core Entities

- **Protocol** (e.g., Uniswap, Aave, Lido)
- **Blockchain** (e.g., Ethereum, Polygon, Arbitrum)
- **Category** (e.g., DEX, Lending, Liquid Staking)
- **Token** (e.g., ETH, USDC, stETH)
- **Organization/Person** (e.g., founders, auditors, venture capital firms)

Relationships

- DEPLOYED_ON (Protocol → Blockchain)
- BELONGS_TO_CATEGORY (Protocol → Category)
- SUPPORTS_TOKEN (Protocol → Token)

- FOUNDED_BY (Protocol → Person/Organization)
- PARTNERS_WITH (Protocol ↔ Protocol/Organization)
- AUDITED_BY (Protocol → Organization)
- HAS_TVL (Protocol → Value)
- BACKED_BY (Protocol → VC/Organization)
- MIGRATED_TO (Protocol → Blockchain)

Data Requirements

- Basic metadata: launch date, website, and blockchain
- Financial metrics: TVL, trading volume, yield
- Temporal data: historical TVL and launch year
- Categorical data: protocol type and category
- Relational data: partnerships, audits, and shared assets

This structure will enable the LLM to perform context-aware reasoning over the DeFi knowledge graph using the GraphRAG paradigm.