COMP39/9900 Computer Science/IT Capstone Project
School of Computer Science and Engineering, UNSW

Project Number: P38

Project Title: ESG Metrics Recommender Using Stardog Knowledge Graph

Project Clients: Prof. Fethi, Mingqin Yu, Dr. Basem Suleiman.

Project Specializations: Big data analytics and visualization; Artificial Intelligence (Machine/Deep Learning, NLP); Software development.

Number of groups: 5

Main contact: Mingqin Yu

Background:

The growing significance of Environmental, Social, and Governance (ESG) metrics in investment decision-making, corporate reporting, and risk management highlights the necessity for a comprehensive solution. This solution must effectively capture, represent, and analyze the complex landscape of ESG data. This project aims to implement a knowledge graph in the ESG metric domain, assisting corporations in cataloging and navigating ESG reporting requirements, standards, and associated data. The Stardog platform will be utilized to build and query the knowledge graph through its designer and explorer functionalities.

Project Requirements:

This project will develop a knowledge graph based on industry material sustainability issues and ESG Metrics. The primary focus will be on employing state-of-the-art Natural Language Processing (NLP) methods and algorithms to match requirements from 11 industries and 77 subindustries with their corresponding metrics from the Clarity AI ESG Risk data source. These metrics will be visualized via Stardog Cloud.

Specifically, the project requires:

User side:

1. Industry-Subindustries Queries: Users can query to see the available industries and subindustries (Already Provided by the customer).

2. Subindustry-Specific Metric Queries: Users can query specific subindustries, and the knowledge graph will provide a list of metrics along with other relevant attributes. The Financial Industry and its subindustries must be included, ensuring robust data retrieval.

3. Company Queries: Users can check which companies are involved for specific metrics.

4. According to matched metrics, User can further check the available data points and their reporting companies from Clarity AI.

System side:

1. System can use NLP to extract the requirement metrics from 77 subindustries PDF files and generate a CSV file or preferred format. The file should include Industry, Subindustry, Requirement Metrics, RMetrics Description, RMetrict, and other relevant

attributes.

2. System can use NLP to link and match the Metrics from Clarity AI data source with requirement metrics and generate a CSV file or preferred format. The file should include Industry, Subindustry, Requirement Metrics, RMetrics Description, RMetrict, Matched Metrics, and other relevant attributes.

Project Scope:

The project will use:

•   Industry requirement metrics from 77 PDF files.

•   Clarity metrics from the Clarity AI metrics introduction CSV file.

•   Clarity AI ESG Risk—Raw Data on Corporations database from EUROFIDAI, including 20 CSV files covering 105 metrics such as total CO_2 emissions, water withdrawal, CEO compensation, and board diversity.

The expectation for system and students:

•   Explain the methods or algorithms used for matching metrics from required industries and existing metrics from data sources.

•   Identify metrics that align with the given ESG industry and subindustries.

•   Visualize the relationship entities through queries on the Stardog platform.

•   Detail the required metrics in each subindustry and the matched metrics from Clarity AI.

Innovative Features:

Design two additional features utilizing queries from the knowledge graph to enhance user experience by providing comprehensive query results. These features maximize the use of the data source for ESG subindustries and metrics, offering deeper insights and more robust analysis capabilities.

Required Knowledge and skills:

•   Natural Language Processing (NLP)

•   Knowledge Graph Development

•   Big Data Analytics and Visualization.

Expected outcomes/deliverables:

•   Documented source code

•   CSV files

•   Shared endpoint (with Owner Access Level to all databases)

•   Final recorded demo

•   Comprehensive project documentation covering knowledge graph models, mapping in corresponding CSV files, explanations, user workflows, and user guides

•   A section on challenges encountered.

Supervision:

Mingqin Yu

Additional resources:

Sample workflows and examples are available.

Collaboration through: ESG teams channel

By participating in this project, you will receive support and feedback from the client and agree to share your work with the client while adhering to confidentiality requirements, including not sharing any work publicly without permission.