

Project Proposal

- **Team Members:** Soojung Choi, Tejas Sujit Bharambe

- **Domain:** We plan to build a knowledge graph about international trade that occur between all the countries in the world. KG in this domain is necessary as the world is connected in the most complex way and with world trade happening 24/7, it generates a lot of data. Hence, we need a smart way to organize and manipulate the data to get answers. This knowledge graph will include primary entities like country, continent, product (ex. *meat of bovine animals*), product sub-category (eg. *meat*), and product main category (eg. *animal products*). The entities would have the primary information like the trading year, import value, export value, and some supporting information like bilateral FTAs between countries, product complexity, and rank, mean tariff, GDP, GDP growth rate, Gini coefficient, human development index, ease-of-doing-business rank, gross external debt, and population. The ideology is to create a KG that will help people discover the dependency of all the countries on each other for all the resources that they utilize. Primarily, this KG will help users explore yearly trade information broken down at either the product, product - country, or country-country level. We also plan to add some interesting elements like how the bilateral FTA between countries impacts trade, how much trade high GDP countries conduct with low GDP countries, what are the most complex products that are traded, etc.

- **Datasets:** This KG will contain the data crawled from the following sources:

- <https://oec.world/> (unstructured, World trade data)
- <https://en.wikipedia.org> (structured, infobox for country-related economic information) (structured, FTA list)
- <https://www.worldometers.info/> (unstructured, yearly population of the country)

There is no existing ontology that fits perfectly for our use case, so we will design a custom ontology that captures all the information about the entities and the relationships between them. Representing the information on KG is a big challenge here which has been outlined below in technical challenges.

- **Technical Challenges:** The major technical challenge is creating an efficient knowledge graph/ontology as we have the following cases:

1. **China** imported **soybeans** from the **US** worth **\$1B** in **2020**
2. **US** imported **soybeans** from **China** worth **\$0.5B** in **2020**
3. **China** imported **soybeans** from the **US** worth **\$1B** in **2019**
4. **US** imported **soybeans** from **China** worth **\$0.5B** in **2019**
5. **India** also imported **soybeans** from **China** worth **\$0.2B** in **2020**
6. **China** also imported **soybeans** from **India** worth **\$0.25B** in **2020**

Hence, we would need a way to handle these bidirectional and complex relationships. Additionally, evaluating the quality of KG is equally challenging. The complexity of analyzing and visualizing the KG is also dependent on the quality of this KG. We plan to use SPARQL/Neo4j queries to extract the relevant information, display KG with Neo4j, and build visualizations using libraries supported by streamlit (<https://docs.streamlit.io/library/api-reference/charts>). The overall idea is to create a UI to allow the users to interact with our application and help them search the desired trade-related information effectively. (Here, we would also have to deal with entity resolution for countries and products). Finally, we will incorporate the predicted trade value (either export/import) of products for the next 3 years using time series-based forecasting models. To evaluate the accuracy, we will perform a qualitative user study by asking our friends to use the interface and hold out data from the two most recent years to evaluate our predictive models.