Project Proposal Guidelines:

Your **1-page** project proposal is due by 11:59pm Sunday,, 2/28/21 via Blackboard. Each group (of two students) should submit a single proposal. The goal of your project is to solve an interesting problem, demonstrate your mastery of building knowledge graphs, and have fun. Your proposal should include three sections:

1. Project Domain & Goals
2. Datasets & Representation
3. Technical Challenge

Project Domain & Goals

This section should clearly explain what your project is about. What sort of information will your KG contain? Why do we need a KG for this domain? What problems will you be able to solve if you build this KG?

Datasets

Where will the data for your project come from? Be specific! Include URLs for the sources you plan to use. How will you represent the knowledge in your KG? Are there existing ontologies that you will use or will you have to develop your own representation?

As a guideline, we expect your project to use at least three distinct sources of data that contain 20,000 pages total. The project should also use a structured source (such as a table, spreadsheet, or knowledge graph) with 500 records. Your KG should have at least 8 semantic types. Part of your project should include some KG analysis (using your KG to answer an interesting data science question) and you should plan to have some sort of GUI to show off your KG.

Technical Challenge

What is the hard technical problem you will solve? Why is it hard? How does it go beyond the material you have seen in lecture? How will you know if you succeed in solving it?

Example

- Domain: We plan to build a knowledge graph about board games. The knowledge graph will include information about the board game genre, number of players, year of publication, playing time, designers, distribution, game mechanics, themes, rules, review scores, and photos. This knowledge graph will be useful because it will help people discover games that they may not know about and also help them organize games they know, for example to find a game from their collection to play given the time and players they have. We plan to use this knowledge graph to visualize how board games have changed over time, including the designers, themes,

and mechanics that have been popular each year, as well as potentially predicting the next "hot" game.

- Datasets: The knowledge graph will contain data mined from structured sources like [BoardGameGeek](#) and blogs about board games such as [Shut Up & Sit Down](#) and [The Dice Tower](#). There is no existing ontology for board games, so we will design a custom ontology that captures the same concepts as the BoardGameGeek structured data (games, designers, themes)
- Technical Challenge: The project will solve novel technical challenges in analyzing and visualizing knowledge graphs. We will use SPARQL queries on our knowledge graph to extract interesting trends over time, such as the prevalence of themes, mechanics, designers, etc. in new board games. Then we will produce timeline and graph visualizations of these trends using the VisJS library. We will also use these trends as well as major board game awards to build a predictive model of themes, mechanics, or trends that are becoming popular using methods such as probablistic graphical models such as PSL and knowledge graph embeddings. We will perform a qualitative user study by asking our friends to use the interface and hold out data from the three most recent years to evaluate our predictive models.