Project Proposal: Higher Education

Ryan Ball, Aaron Yang, Kevin Tran

September 25, 2020

Project Domain and Goals

This project will focus on looking at higher education in the United States. The knowledge graph will contain information pertaining to universities across the nations such as major and programs, rankings, statistics such as acceptance rates, tuition, etc, It will also contain information about the campus such as its colors, founding, location, founder, logo image, student life, etc. A knowledge graph is needed for this domain because millions of students across the world have a hard time sifting through which college they want to go to. Students tend to go for more popular schools but aren't aware of other schools that may have a better program, offer better financial support, or have a good ranking. By incorporating knowledge on school statistics, university profiles, and other factors, a person can get a better overview of colleges that they might want to go to. Other problems include getting a more accurate depiction of acceptance rate and rankings, finding majors and programs that fit one's interest, and determine aspects of student life.

Data sets

The data for the project will come from the following sources:
1) US News College Ranking: https://www.usnews.com/best-colleges
2) College Confidential: https://www.collegeconfidential.com/schools
3) Niche: https://www.niche.com/colleges/search/best-colleges/
4) DBpedia 5) Wikipedia: https://https://en.wikipedia.org/wiki/
6) US Department of Education: https://collegescorecard.ed.gov/school/
7) World University Ranking: https://www.timeshighereducation.com/world-university-rankings

The knowledge will be represented as the school for entity and the links to be all the various information about the school. Schools may share the same program type, be classified as a 4-year institution, acceptance rate, campus setting, etc. There are existing ontologies for academia but we may need to modify it with our own to accommodate different information.

Technical Challenge

The project will solve the technical challenge of consolidating different factors from different sources to help a person determine if the college is the best fit. It is difficult because across all the sources there different metrics to score universities, the problem with entity linking of majors and programs, similar school names. It goes beyond the material seen in lectures because it will utilize the knowledge graph to recommend and assist in choosing the best college. The success of the project will be based on it's usefulness, being able to give out accurate information, and being able to visualize different statistics for this aggregated data.