

Group Members: Gabriel Centeno, Anirudh Kaushik, Ethan March

Group Name: CentenoKaushikMarch

Project Description:

While the general SQL language is pretty simple and does not require a lot of programming “know-how” (e.g. OOP, algorithms), some individuals who want to work with large data sets may not feel comfortable working directly with the language and mySQL Workbench or some other database design tool. This is the ultimate goal of a database management system (DBMS). We want to build a DBMS that abstracts the programming language of SQL out of the user interface, and instead replaces it with something every end user should know: English. Natural Language Processing is a fast-growing field of computer science research, and many popular tools utilize its technology to bring information closer to users (think of things like Siri, Google Assistant, Amazon Alexa).

We are looking to design a DBMS which can translate simple English requests into SQL queries. This DBMS will have two main components: a natural language processor that turns an English phrase into a SQL statement (either with the DDL or the DML), and a script that acts on a database. In order to retrieve the intended SQL query from the user’s input, will use IBM Bluemix and Watson’s Natural Language Classifier. Once we have interpreted the input and created the SQL query, we will have a Python script that interacts with the database in order to run the query against it.

The dataset we want to work with is school data. This includes students and the classes they want to take during the semester, teachers and the classes they can and can’t teach, class size, course prerequisites, schedule conflicts, and things of that nature. This data set is relatively standard and lends itself to the modeling we have done in class, ensuring that we will do a good job in structuring our database. Importantly, this data must be consistent for a variety of constraints that are not apparent directly from the data itself (scheduling conflicts, prerequisites, etc). This means that manipulating the data in this set is not as simple as a SQL command completing and making one change.

Since the IBM tools are the ones we are least familiar with, we believe most of our time will be spent working with these technologies. As a result, our goals will progress from having the user provide input in some simple structured format (almost a SQL command, but in a “sentence”) to a point where the user should not recognize that they are working with a SQL tool. This end goal, while inspirational, is not one we believe we could achieve in the span of this semester, but we will always work towards designing a DBMS that can accomplish this.

Languages:

Python, Javascript, HTML

Hosted on Bluemix with SQL storage

Using IBM Watson Natural Language Classifier

Interest:

We are interested in working with this kind of data because Anirudh particularly wanted to construct a scheduling software

Anirudh: In high school, it was often the case that students' schedules would be released very late, in some cases even a day or two after school started. My friends and I would often wonder how this could happen, especially considering the counselors had all summer to arrange this, and the fact that there must be some standard software that schools use to devise schedules. I want to see exactly how complex this kind of software is and, in particular, if my guidance counselors have any defense for their lateness with my schedules.

Request for three person group:

We would like to work in a three person group because none of us have ever used these cloud services before and the scope of what we are trying to do is very large; we also have other coursework and extracurriculars that will take time away from completing this project. Ethan doesn't have any Javascript experience which Gabriel does have, while Gabriel has never used Python and Ethan has. Anirudh has dabbled a bit in both but has not done a large project like this in Python. We believe that we will learn a lot in this diverse group and be able to create a very cool project that we are genuinely interested in.