

## CS411 PT1 Stage 4 Project Report

Link to video:

<https://youtu.be/YpnAxdOFpxg>

**Please list out changes in the directions of your project if the final project is different from your original proposal (based on your stage 1 proposal submission).**

There is no significant change to the main direction of this project. It still conforms to the concept we initially put forward in the stage 1 proposal

**Discuss what you think your application achieved or failed to achieve regarding its usefulness.**

Usefulness: Successfully enabled users to keep a running list of medications; Display to users interactions of taking different medications with a known type of medication

**Discuss if you changed the schema or source of the data for your application.**

Our initial schema as presented in the original stage 2 submission ended up being infeasible due to the different ways in which drugs were uniquely identified between datasets (more in the table implementation section below). As for the sources of the data, we reduced the number of data sources to 3: the DDID database for drug-food interactions, RxNorm database for a list of drugs and similarity relationships between drugs, and twoSides database for a comprehensive list of interaction pairs between drugs as these three together amounted to several GB of data alone.

**Discuss what you change to your ER diagram and/or your table implementations. What are some differences between the original design and the final design? Why? What do you think is a more suitable design?**

The original schema design submitted for stage 2 had separate tables for drugs and foods, but due to us deciding to use the RxNorm standard for indexes, we would have had to reformat all the unique foods to use the same standard, and at that point, it was easier to simply merge the two tables into one. Also, there were some problems related to using separate data sources; the twoSides interaction table used only RXCUI identifiers, which is essentially an ID for a group of closely related RXAUI identifiers. However, it turned out that the RxNorm tables used RXAUIs to uniquely identify a drug, and it was not possible to efficiently reconcile the two different ways of storing data, so we simply opted to solely use RXCUIs as identifiers.

Also, the way that RXCUIs are structured in the original data made working with ids incredibly difficult. As an example, a drug like acetaminophen has a RXCUI of 161, but there are over a hundred related RXCUIs all standing for some variation of acetaminophen/branded version like Tylenol/generics like paracetamol and then reconciling all of these on a per drug basis (when there are over 400,000 unique RXCUIs and over a million RXAUIs) virtually impossible.

### **Discuss what functionalities you added or removed. Why?**

By enabling users to manage a list of medications, we have incorporated the functionalities of four aforementioned basic functionalities. This significantly reduces the burden individual users have to bear by visualising their choices.

### **Explain how you think your advanced database programs complement your application.**

By enabling users to keep track of the list of medication they take, our application weaved together four basic functionalities to help users organize their query results into a final decision of medications.

### **Each team member should describe one technical challenge that the team encountered. This should be sufficiently detailed such that another future team could use this as helpful advice if they were to start a similar project or where to maintain your project.**

- Daniel: A major challenge was the sheer scale of the datasets we had; it made working with a minimally provisioned google cloud instance all but impossible with the final database ending up being about 6.24 GB and not being able to be imported into the GCP instance at all. The size of the database did make writing efficient queries very important, but the downside was making managing local versions of the database quite difficult between group members and generally made collaboration difficult.
- Pranjali: Generating test data was somewhat challenging due to foreign key constraints. Since you had to reference valid foreign keys, you had to make sure you were using real values. I did this manually for some tables, but I think automating random entry generation through a script would have sped things up and created a lot more data.
- Hank: Loading the entire sql dump while making changes to certain details of the script in order to get it going is harder than expected. The dependency between different tables made it challenging.

### **Are there other things that changed comparing the final application with the original proposal?**

The original proposal suggests using imported medical records for each user, but this turned out to be infeasible due to medical privacy laws, so we opted for having just checking for possible interactions and listing potential medical conditions which could cause increased risks.

### **Describe future work that you think, other than the interface, that the application can improve on**

We hope to load in the price of medicine as well as its availability. This potential application can be used in Medicaid plans as patients can be informed of different choices of treatment as well as the cost. It can also help government agencies review certain spendings.

Additionally, working on linking the various branded versions of drugs together when adding to medication sets and then viewing the interacting drugs can significantly improve the intuitiveness of the application.

**Describe the final division of labor and how well you managed teamwork.**

Group coordination was done through a discord server. Work division:

- Daniel: data cleaning/preparation, SQL database implementation, database integration with backend, frontend UI construction, API integration
- Jason: Frontend UI construction, writing advanced SQL queries
- Pranjal: writing advanced SQL queries, frontend UI construction, API integration
- Hank: Documentation, video recording