Foodomics Database Metabolite Calculator

Gregory Scott Wilkinson

CIS4914, Senior Project

Department of CISE

University of Florida

Advisor: Dr. Peggy R. Borum, *email*: prb@ufl.edu

Abstract

In Dr. Borum's lab, the foodomics database contains all relevant information about the foods needed to treat patients. In the past, this database has been mostly reliant on the information that is provided on the nutrition facts label. My project, the Metabolite Calculator, is a subsection of the foodomics database that contains information regarding the individual fatty acids, amino acids, and sugars/carbs that make up the foods in the database. This is achieved by creating "composite generic foods" of similar foods from a database that contains all the required information, such as the USDA's sr28 database, and using this information to create an educated guess for the makeup of brand name foods in the foodomics database. My project also entailed creating a frontend for clinical members to be able to interact with the database to add, edit, or delete food profiles as needed. This was done using Flask and Python for the frontend and PostgreSQL to connect to the database.

Introduction

The goal of Dr. Borum's lab is to treat patients that suffer from seizures by putting them on a precision keto diet. As more information is obtained about how different elements of diets can affect patients, more specific information about the food given to them is needed to improve treatment. For my senior project, my plan was to create a section of the foodomics database to help clinical members within Dr. Borum's lab obtain and organize information regarding foods and oils that are used to treat patients. I am doing this project because I have worked with databases and Python in the past and I felt that this would be a good opportunity to gain practical experience as well as being able to help patients. While I am the only one working on the Metabolite Calculator, it will be integrated into a single application that involves projects from other computer science students. These other projects will pull data from the Metabolite Calculator section of the foodomics database for their own calculations. During this project, I faced some problems such as my limited knowledge of food science as well as considering possible improvements that will need to be made in the future as treatments for patients evolve over time.

Problem Domain

As treatments for patients keeps evolving, more information about the food and oils that is prescribed to them needs to be available. While the exact information is difficult to obtain, an estimate can be calculated to help clinical members make more effective meals for treatments. This process for calculating information about a food's profile can take a considerable amount of time to do by hand. In response to this, the Metabolite Calculator aims to automate these calculations and add all necessary information to the foodomics database in order to make this process more efficient.

Previous Work

For my project, there is very little previous work from outside sources. My project within Dr. Borum's lab is, to the best of my knowledge, the first of its kind. While there are no other examples to the Metabolite Calculator, many details about how it could be created were explained in a thesis paper by Lujia Yang. The focus of Lujia Yang's thesis the creation of the entire foodomics database and how the database can be applied to help patients. Using this thesis paper as a blueprint, it was possible to create the Metabolite Calculator without requiring an extensive knowledge of food science.

The reason Lujia Yang's thesis is so important to my project is the methods and information it provides about how the Metabolite Calculator is supposed to function. The Metabolite Calculator needs to be divided into three parts to create profiles of fatty acids, amino acids, and the carbohydrates/sugars that makeup food. The data used to create profiles in each of these sections is taken from the USDA sr28 database [1]. In the past and as of the time this paper was written, this database is still being used, however it is a legacy database. Given the data in the USDA sr28 database is not always complete for each of the three sections for every food, calculations done for each section must be separate [1] [2]. Lujia Yang's thesis details how this data can be taken and used to create a composite food in order to estimate the molecular profiles of the brand name foods.

Technical Approach

In order to estimate the molecular profile for brand name foods, a composite food must be made first. The way my project is designed, a user must start by finding food entries that are

similar in name from the USDA sr28 database (or from a similar database, all that is required is to have the necessary molecular information). Taking this information, the user can create a composite fat profile, composite protein profile, or a composite carbohydrate profile [1]. From here, we will assume the user is trying to make a composite generic fat profile for apples in order to demonstrate how composites are created. A user would then enter the total fat, saturated fat, monounsaturated fat, polyunsaturated fat, and trans-fat per 100g of food from each apple entry in the USDA sr28 database that was deemed relevant to create a composite. The Metabolite Calculator will calculate the standard deviation and mean for each of these values across the entries and create a range from the mean ± the standard deviation. Any entries with values that are outside of this range are discarded, as they are too different to be considered for the composite. Of the remaining entries, the user will enter all the individual fatty acid values and the Metabolite calculator will create a profile that is the average of all the remaining entries. From here an id value is given to the profile and it is added to the fatty acid composite generic foods table in the database. To create the protein and carbohydrate composite profiles, the process is the same except values related to amino acids or carbohydrates/sugars are used instead of fatty acids.

Fatty_Acid_ProfileCGF		
	Name	VARCHAR(100)
PK	CGFID	VARCHAR(100)
	TotalFat	Decimal
	TotalSat	Decimal
	Butyric	Decimal
	Caproic	Decimal
	Caprylic	Decimal
	Capric	Decimal
	Lauric	Decimal
	Tridecanoic	Decimal
	Myristic	Decimal
	Pentadecylic	Decimal
	Palmitic	Decimal
	Margaric	Decimal
	Stearic	Decimal
	Arachidic	Decimal
	Behenic	Decimal
	Lignoceric	Decimal
	TotalMono	Decimal
	Myristoleic	Decimal
	Pentadecenoic	Decimal
	Gaidic	Decimal
	Palmitoleic	Decimal
	Sapienic	Decimal
	Heptadecenoic	Decimal
	Margaroleic	Decimal
	Ricinoleic	Decimal
	Oleic	Decimal
	Elaidic	Decimal
	Gadoleic	Decimal
	Docosenoic	Decimal
	Cetoleic	Decimal
	Brassidic	Decimal

Figure 1. Sample of the ER diagram for the fatty acid composite generic foods table. The primary key for each of the composite generic food tables is the "CGFID". While two fatty acid composites cannot share the same id number, this id number can be used to identify the composites for a food's amino acid or carbohydrate composite profile. For example, if the id for red apples were "CID001" on the fatty acid table, then the id for red apples on the amino acid composite and carbohydrate composite tables would

also be "CID001". The other composite tables have similar makeups to the one shown above, involving a name, CGFID, all values needed for the type of profile, and a date that the profile is created. The tables for the brand name profiles match these exactly except have an "NDID" value rather than a "CGFID" value.

Once there is a composite for a food, it then is possible to estimate the molecular components of brand name foods. Using the example from before, assume the user is trying to make a fat profile for specific brand of apples. Using information from the nutrition facts label, it is possible to scale a composite's information to meet that of the brand name. The user is prompted to put the total fat, saturated fat, monounsaturated fat, polyunsaturated fat, and trans-fat from the nutrition facts label. Since nutrition facts labels use serving size, all this data needs to be scaled up to 100g of food in order to match the data from the composite. This is done to know the appropriate scale factor for each of the individual fatty acids. For all fatty acids that fall under the category of saturated fat, the value from the composite will be multiplied by the value the user input for saturated fat, and then it will be divided by the composite value of saturated fat. This process is the most accurate way to estimate the individual fatty acids, however the limited information that is given on the nutrition facts label makes this process not always possible. Often, saturated, monounsaturated, and polyunsaturated fat is not listed on the nutrition facts label. To accommodate for this, for any value that is not available, it will use total fat for scaling. Once this is done, the new profile for the brand name apples is tagged with its NDID and saved to the fatty acid profile table in the database. Like the composites, the process for creating amino acid and carbohydrate profiles is like that of fatty acid profiles.

Add Fatty Acid Profile

Name

Food Name

NDID

CGFID

Total Fat

g

Total Saturated Fat

Optional g

Total Mono Fat

Optional g

Total Poly Fat

Optional g

Total Trans Fat

Optional g

Create Profile

Figure 2. This is the input form to add a new fatty acid profile for a brand name food. This form prompts the user to set the name and NDID of the new profile as well as enter the CGFID of the composite being used to create the profile. After this the total fat is a required value to create the profile, while all other values are optional for if they are available. The input forms for the amino acid profiles and carbohydrate profiles use a similar as described above. Once the values are entered, the composite that was chosen is then scaled in the backend and the new brand name food profile is added to the appropriate table in the database.

Results

Once all the calculations were programmed and the frontend was created, the Metabolite Calculator was used to create profiles for some foods in the foodomics database. Each of the composite and brand name tables have pages to view the contents of the database. Using the profiles that were created, it was possible to see how the calculations generated the estimated values for the molecular makeup of the foods and oils. While the Metabolite Calculator works exactly as intended, if the data that was used to create the composites is incomplete, then this will affect the final profile. The USDA sr28 database is a legacy database and many entries may have incomplete information. Due to this, when viewing data in this section of the database, a zero does not necessarily indicate there is 0g of that value, but instead means that none of that value was detected. For instance in a fatty acid profile, if the value for butyric acid is zero, that means that the profile may contain some butyric acid but the amount is unknown.
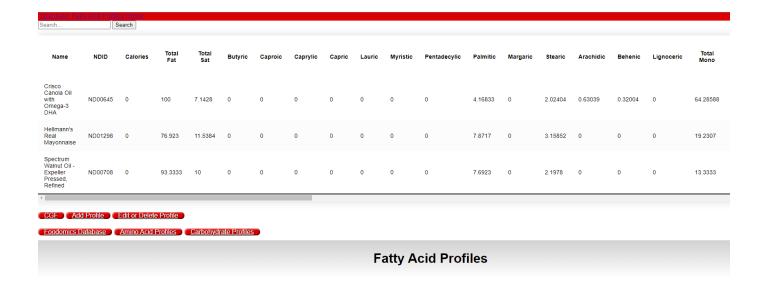

Name	NDID	Calories	Total Fat	Total Sat	Butyric	Caproic	Caprylic	Capric	Lauric	Myristic	Pentadecylic	Palmitic	Margaric	Stearic	Arachidic	Behenic	Lignoceric	Total Mono
Crisco Canola Oil with Omega-3 DHA	ND00645	0	100	7.1428	0	0	0	0	0	0	0	4.16833	0	2.02404	0.63039	0.32004	0	64.28588
Hellmann's Real Mayonnaise	ND01298	0	76.923	11.5384	0	0	0	0	0	0	0	7.8717	0	3.15852	0	0	0	19.2307
Spectrum Walnut Oil - Expeller Pressed, Refined	ND00708	0	93.3333	10	0	0	0	0	0	0	0	7.6923	0	2.1978	0	0	0	13.3333

CGF   Add Profile   Edit or Delete Profile

Foodomics Database   Amino Acid Profiles   Carbohydrate Profiles

Fatty Acid Profiles

Figure 3. This picture shows the fully calculated values of some brand name foods in the foodomics database. Each of these has values based off the composite that was made for that type of food. For instance, the top entry, "Crisco Canola Oil with Omega-3 DHA" was made from a composite canola oil profile. This brand name canola oil, given that there was 100g of it, would have 7.1428g of saturated fat. The Metabolite Calculator took the composite profile for canola oil and scaled the individual fatty acids to estimate what was in the brand name canola oil. It can be confirmed that the calculations

worked in estimating the individual fatty acids as all the saturated fatty acids adds up to 7.1428g. If the sum is lower than this, it is likely due to the composite being made from data that is incomplete to some extent.

Conclusion

Over the course of this project, I felt that I gained a lot of experience working with a real-world problem that could be solved through my experience and knowledge in software engineering. I feel as though my solution can be a great help to current and future treatment for patients done by Dr. Broum's lab. The Metabolite Calculator aims to solve a lack of necessary information by calculating estimates of molecular profiles of nutrients in foods. That data that is calculated is, as of now, the most accurate way of quickly determining the makeup of brand name foods beyond what is given on the nutrition facts label. While I believe the calculator takes the current process and makes it more efficient, there are still many manual processes that could be sped up in the future, such as creating the composite generic foods. This process may be possible to be automated further if a new database replaces the USDA sr28 and information can be pulled directly from the new database to make entering in data faster. As of now, the manual process works and is adaptable for wherever source data is pulled from. Another way the Metabolite Calculator can be improved in the future is when creating new brand name profiles, only one composite profile can be used. For certain types of foods, such as a mixed bag of vegetables, it may be necessary to pull data from multiple composites in order to create the brand name profile.

Acknowledgement

I would like to thank Dr. Borum as well as all the clinical members for their guidance and encouragement throughout the project. I would also like to thank the other computer science members working within the lab for any technical help that was needed.

References

[1] L. Yang, *FOODOMICS DATABASE: A NEW TOOL FOR ESTIMATING MOLECULAR,* Gainesville: University of Florida, 2018.

[2] USDA, "SR28," May 2016. [Online]. Available: https://www.ars.usda.gov/northeast-area/beltsville-md-bhnrc/beltsville-human-nutrition-research-center/methods-and-application-of-food-composition-laboratory/mafcl-site-pages/sr11-sr28/. [Accessed November 2021].

[3] "Flask and Databases," 2020. [Online]. Available: https://python-adv-web-apps.readthedocs.io/en/latest/flask_db1.html. [Accessed 2021].

Bibliography

I am Gregory Scott Wilkinson, and I was born on February 12, 1999. I completed high school at East Lake high school and am currently completing my bachelor's degree in Computer Science at the University of Florida. I expect to graduate on December 18, 2021. In the past, I have interned for Amgen for two summers working in Data Loss Protection. After graduation I will be returning to work with a full time position in software engineering at Amgen.