GLUKOZA -- design, explanations, and rationale of a diabetes management conversational agent

Solution for the **Alexa Diabetes Challenge**

Anton Antonov
Clearsense LLC
May 2017

Introduction

This document provides general description and design document references of a proposed solution named “***GLUKOZA***” for the “Alexa Diabetes Challenge” competition (https://www.alexadiabeteschallenge.com).

Assumptions

We use the following assumptions for the design of our solution.

- There are detailed, comprehensive solutions of diabetes management that exist, like “AACE/ACE comprehensive type 2 diabetes management algorithm”, [14].
- Most users with diabetes would have type 2 diabetes (≈ 90 %).
- “Type 2 diabetes is usually first treated by increasing physical activity, and eliminating saturated fat and reducing sugar and carbohydrate intake with a goal of losing weight. These can restore insulin sensitivity even when the weight loss is modest, for example around 5 kg (10 to 15 lb), most especially when it is in abdominal fat deposits. Diets that are very low in saturated fats have been claimed to reverse insulin resistance.” See [15].

Scope of the proposed solution

1. GLUKOZA has functionalities for two types of users:

1) a type 2 diabetic user that wants to manage and reverse his condition, and
2) a researcher that wants to do statistical studies of the behavior of many diabetics users.

2. GLUKOZA’s diabetics management user functionalities:

- facilitate following medicine prescriptions,
- facilitate the inter-operation with dedicated diabetes management apps,
- facilitate regular hospital visits for monitoring of sugar levels,
- facilitate the monitoring of physical activities and food consumption, and
- provide motivation for following plans for weight loss and exercise.

3. GLUKOZA’s statistical studies functionalities:

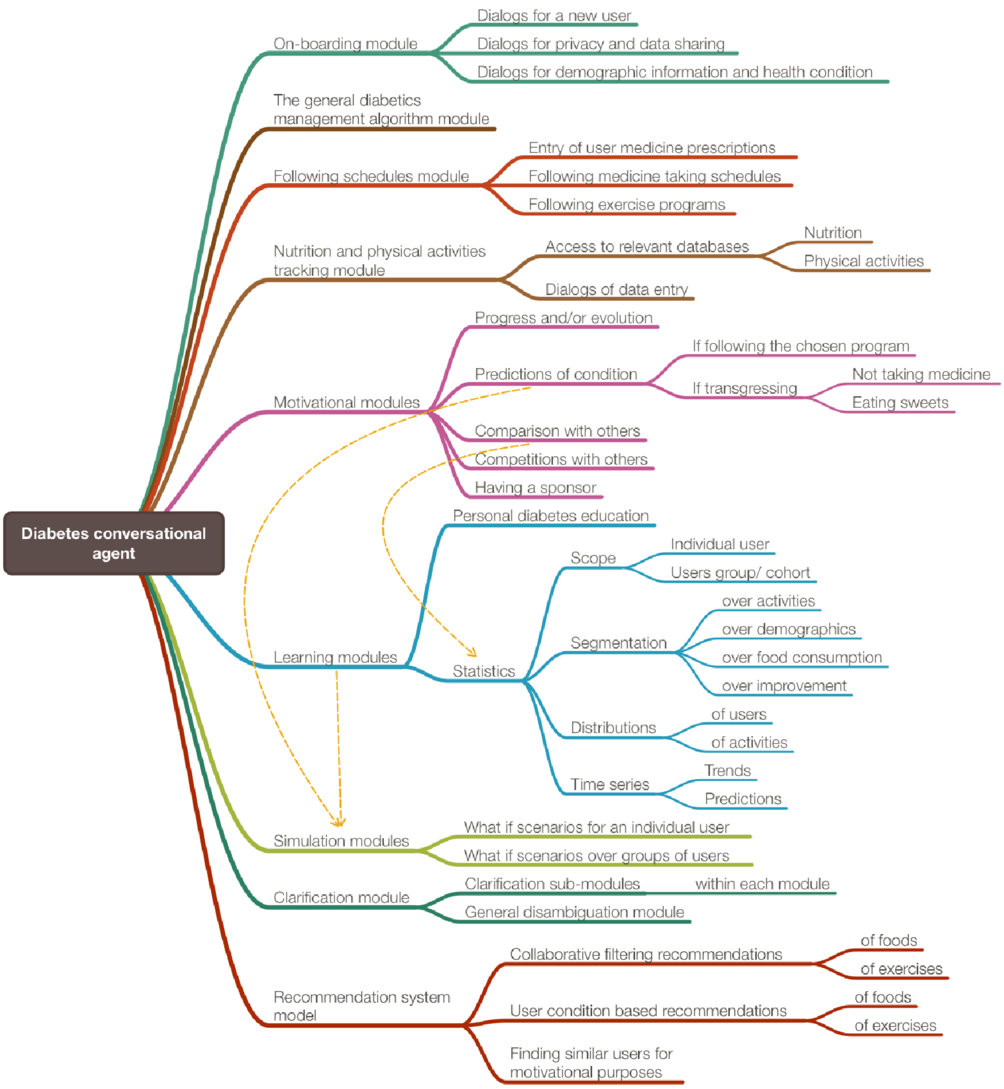
- provide distributions across demographics or behavior types,
- provide time series analysis or trends of adoption, exercise, food consumption, and

- facilitate the experimentation with different "what-if" scenarios.

4. GLUKOZA’s recommendation system
   - provides collaborative filtering recommendations of food and exercises,
   - provides finding of similar users for motivational purposes.

For more a more structured overview see the mind-map "GLUKOZA conversational agent modular design”, [4].

Out[295]=



The name “Glukoza”

The name “Glukoza” was chosen because:
-- it corresponds to the Slavic languages pronunciation of “glucose“;
-- it fits personal assistant names “Alexa”, “Cortana”, “Eliza”;
-- it is (relatively) memorable and easy to pronounce.

Conversational agent design

Functionalities and components

At the brainstorming phase for this project several types of functionalities were identified and considered. They are detailed in the submission document “Morphological analysis of GLUKOZA’s conversational agent elements over functionalities-vs-data breakdown”, [1].

GLUKOZA’s conversational agent components are organized in a way that provides both incremental and modular development and rich and flexible end-user functionalities. See the submission documents “GLUKOZA design”, [3], and “GLUKOZA conversational agent modular design”, [4].

Parsed sentences grammars

At this point we have prototyped several grammars for GLUKOZA using the package FunctionalParsers.m available at GitHub; see [7,8]. Here is an example of natural language sentence parsing:

```
In[234]:= ParseShortest[pUSERCOMMAND][
 ToTokens["i am harry smith and i am type one diabetic"]]

Out[234]= {{{}, {UserName[{harry, smith}], UserDiabet[{{type, one}, diabetic}]}}}
```

Remark: Note that the important information elements in the statement (name, diabetic type) are wrapped into functional expressions. In the interpretation phase the functions are hooked-up with their implementation bodies depending on the interpretation context. More explanations and examples of the underlying methodology, design principles, and software can be found in [6,7,8,9,10].

Several examples of natural language parsing for different components follow. (The corresponding code and more examples can be found in the submission document “GLUKOZA grammars”, [5].)

On-boarding module

Out[264]=

1	command: i am nina parsed: UserName[{nina, {}}] residual: {}
2	command: i am female parsed: UserSex[female] residual: {}
3	command: i am 26 years old parsed: UserAge[26] residual: {}
4	command: my race is asian parsed: UserRace[asian] residual: {}
5	command: my weight is 170 pounds parsed: UserWeight[{170, pounds}] residual: {}
6	command: i am 70 kilograms parsed: UserName[{70, kilograms}] residual: {}
7	command: i am type two diabetic parsed: UserDiabet[{{type, two}, diabetic}] residual: {}
8	command: i am pablo and it seems i am a diabetic now parsed: {UserName[{pablo, {}}], UserDiabet[{{}, diabetic}]} residual: {}

Time series analysis module

Out[293]=

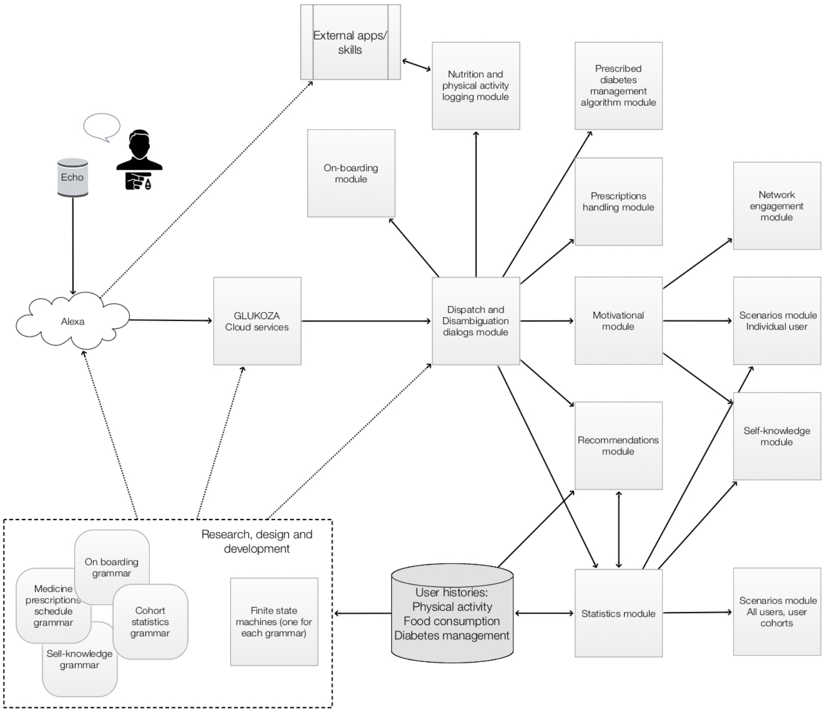
1	command: load the calorie intake of the current user parsed: TSCommand[TSLoadData[TSNutritionSpec[{{}, {TSNutritionElement[calorie], {TSCohortSpec[{the, current, user}], {}}}}]]] residual: {}
2	command: consider the calories of diabetic users in april parsed: TSCommand[TSLoadData[TSNutritionSpec[{{}, {TSNutritionElement[calories], {TSCohortSpec[{diabetic, users}], TSTimeInterval[april]}}}]]] residual: {}
3	command: get the protein intake of diabetic users between new year and now parsed: TSCommand[TSLoadData[TSNutritionSpec[{{}, {TSNutritionElement[protein], {TSCohortSpec[{diabetic, users}], TSTimeInterval[{between, {TimeSpec[{new, year}], {and, TimeSpec[now]}}}]}}}]]] residual: {}
4	command: load the total price of diatex parsed: TSCommand[TSLoadData[TSFinancialData[{TSFinancialElement[{total, price}], {of, TSMedicineSpec[diatex]}}]]] residual: {}
5	command: compute bottom outliers parsed: TSCommand[TSOperateCommand[TSOutliers[{bottom, outliers}]]] residual: {}
6	command: find the top outliers parsed: TSCommand[TSOperateCommand[TSOutliers[{top, outliers}]]] residual: {}

Component diagram

The envisioned design of the different GLUKOZA components is given in the submission diagram “GLUKOZA design”, [3].

Here is small size copy of the diagram:

Out[198]=



Simulations of large number of users usage

The purpose of the simulation

In order to demonstrate the benefits of the “network effect” (https://en.wikipedia.org/wiki/Network_effect) from having a large number of users we are going to simulate GLUKOZA’s adoption and usage by different types of users.

The benefits of the network effect should be seen from the following perspectives:

-- an individual user,
-- researchers and marketeers, and
-- physicians and developers wanting to improve GLUKOZA’s utility.

With the simulation generated data below it should be clear how we can answer user questions like:

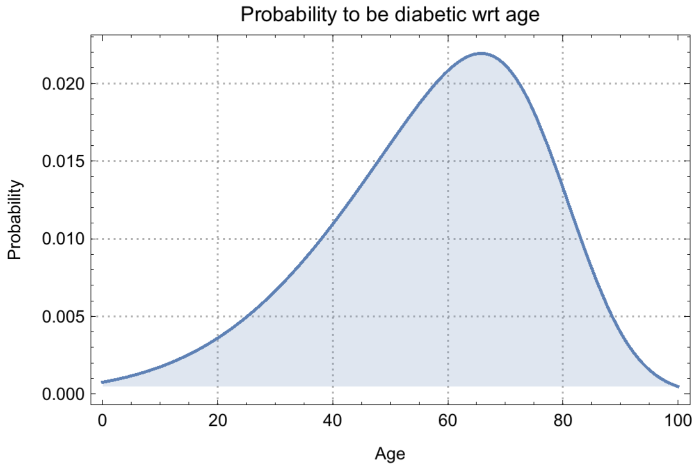
- Where am I at in comparison with others of my demographic?
- How many users with BMI ≥ 30 consume more than 2500 calories a day?
- What is the average protein intake of Florida males with BMI ≥ 28?

Making up users

Finding out diabetic users

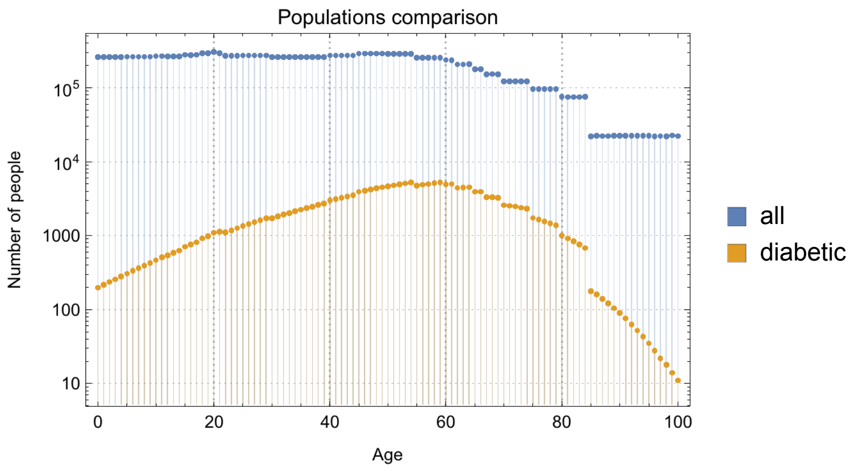
Here is a hand made distribution based on [13] that we are going to use for the probabilities to have type 2 diabetes across ages:

Out[159]=



Next we consider the ages of USA breakdown, [12], from which we can derive the number of people who are diabetic.

Out[176]=



Adding sex, height, weight

Further we proceed by adding sex, height, weight to each user.

1. Assume that the distribution across sexes is approximately uniform : https://www.cdc.gov/diabetes/statistics/prev/national/figbysex.htm .
2. Since we assume users to be adults (18 years old or older) we can take normal distributions for the height of each sex.
   We assume that the average male height is 175.7 cm and the average female height is 161.8 cm. (Using the USA records in this Wikipedia article https://en.wikipedia.org/wiki/List_of_average _human _height _worldwide .)
3. Since we assume that all users are diabetic or pre-diabetic we can simulate user weights by using random Body Mass Index (BMI) numbers in the range [27, 33]. Note, that the BMI distributions across the sexes are assumed the same and uniform.

Adding state

To each user we assign state of residence based on weighed random selection of states. (More populated states have higher probability to be chosen.)

Summary of the simulated user demographic records

At this point we have the following user demographic data summary

Out[194]=

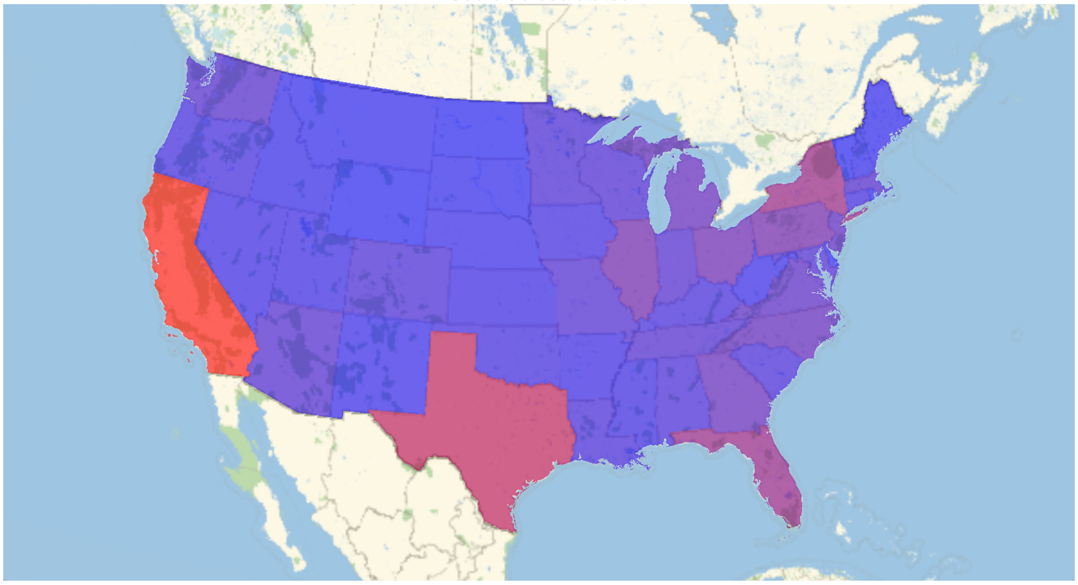
1 id Min 1. 1st Qu 47 721. Mean 95 441.5 Median 95 441.5 3rd Qu 143 162. Max 190 882.	3 sex male 95 794. female 95 088.	5 weight Min 46. 1st Qu 74. Median 84. Mean 86.1246 3rd Qu 96. Max 174.	7 BMI Min 26.7801 1st Qu 28.4799 Mean 30.0001 Median 30.0266 3rd Qu 31.5334 Max 33.2275
2 age Min 18. 1st Qu 43. Mean 52.8541 Median 54. 3rd Qu 63. Max 100.	4 height Min 130. 1st Qu 158. Median 168. Mean 168.753 3rd Qu 179. Max 230.	6 state California 23 141. Texas 16 171. New York 11 955. Florida 11 726. Pennsylvania 7725. Illinois 7600. (Other) 112 564.	

USA map

The following map shows the distribution of GLUKOZA’s users across USA states. The normalized numbers of users per state are used to derive the colors that blend blue and red -- the bluer the less users, the redder the more users.

Out[57]=

Users across states



Obviously, similar plots can be made for different statistics over GLUKOZA’s users. For example, adoption rates, different demographic variables breakdowns, average number of calories per person consumed in a day, and others.

These statistics are computed by voice commands prompts. The statistics outcome can be both displayed (if Amazon Echo Show is around) and/or summaries of them are told to the user.

Making up food histories

At this point we have one or several example food diaries. Each food diary record is for a food consumed at a certain date as a part of a certain meal (e.g. breakfast.) Each record has several

nutritional elements fields (protein, calories, etc.)
Here is an example:

Out[58]=

sodium	total_carbs	calories	sugar	full_food_description
11	14	62	3	Fresh – Corn on the Cob, 0.5 ear, large (7–3/4" to 9" long) yields (143g)
290	15	120	14	Berkeley Farms – Cultured Low Fat Buttermilk–1% Milkfat, 1 cup
53	5	128	2	Body Fortress – Super Advanced Whey Protein Powder– Vanilla, 0.75 Scoop (41g)
55	13	50	10	Smucker's – Sundae Syrup Caramel, 1 Tbsp or 40 grams
22	27	123	5	Fresh – Corn on the Cob, 1 ear, large (7–3/4" to 9" long) yields (143g)
1080	61	510	5	Taco Bell – Beefy Crunch Burrito 07/08/13, 220 g
0	20	128	14	Rt&St – Homemade Chocolate Brownie, 35 g
2666	177	1314	36	Eli's – Homemade Spaghetti–whole Wheat W/ Meat Sauce, 3 serving
155	15	135	10	Cliff Builders – Protein Bar – Chocolate Peanut Butter, 0.5 bar
0	0	100	0	Pedialyte – Pedialyte Frozen Ice Pop, 1 pop
120	2	140	0	Generic – Free Range Brown Egg, 2 egg
190	28	150	4	Sara Lee – Soft & Smooth 35% Whole Grain White Bread, 2 slices (57g)
40	30	160	30	Epicure Selections – Hot Buttered Rum, 4 heaping teaspoons (20g)
70	2	20	0	Northeast – Baby Spinach Raw, 2 cups
5	4	32	3	Sliced Tomato – Roma Tomato (Per Nutritiondata.self.com), 1 tomato

Using the example food diaries we generate a random food diary for each of simulated users.

We can generate the food diaries in several ways that are alternative or complemental to each other.

1. Completely random selection of days for each user.

2. Random days selection based on:
2.1. user demographic (sex, weight, etc.)
2.2. ***and*** day's total calories (or protein, or fat, etc).

3. Random selection of individual records for each user for each type if meal.

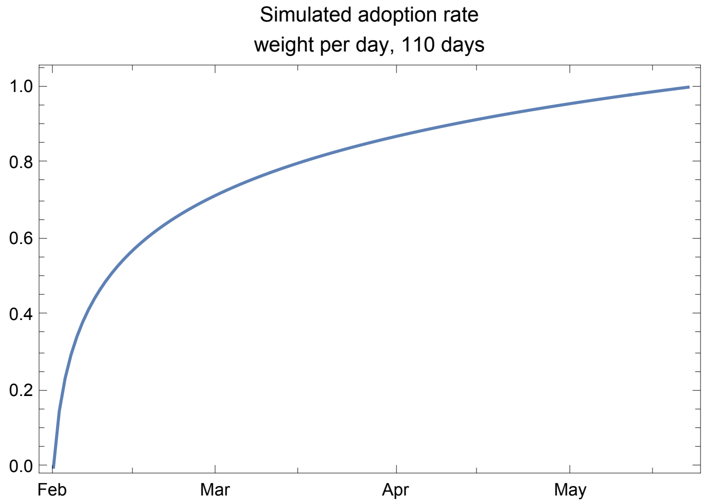
4. Selection based modeling user archetypes:
4.1. consistent users;
4.2. chaotic users;
4.3. users with short app adoption lifespan;
4.4. based diabetic or pre-diabetic condition.

5. Based on app adoption rate modeling:
5.1. social networks effects (e.g. word of mouth);
5.2. recommendations by medical professionals;
5.3. based on app feature enhancements;
5.4. based on marketing activity (advertisements).

Simulated adoption rates

Here is a plot of the simulation adoption rate of GLUKOZA:

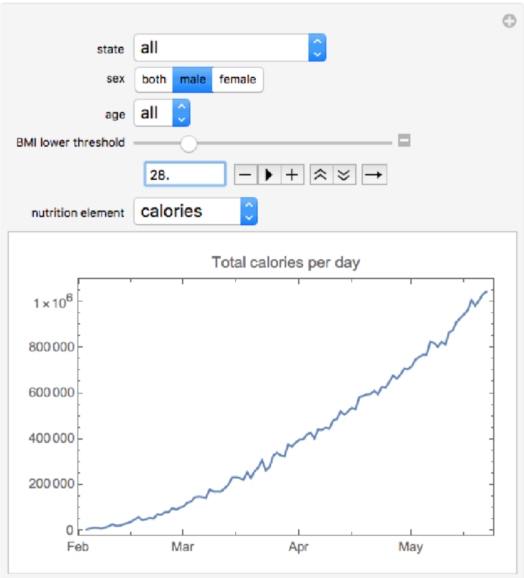
Out[144]=

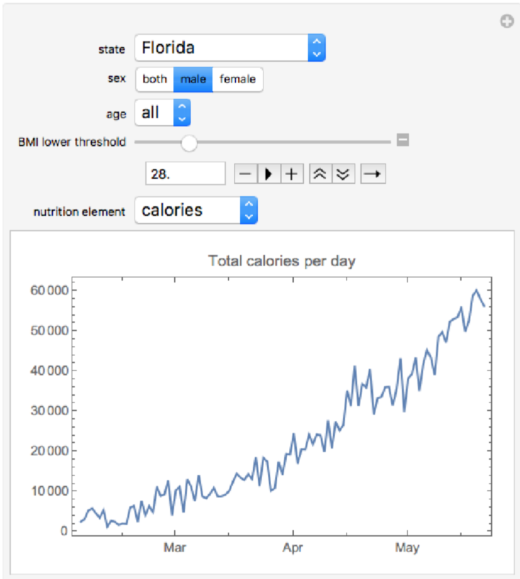


These adoption rate curve is used to randomly select the first GLUKOZA usage day for each user.

Dynamic statistics

Having the simulated data we can make the following dynamic interface snapshots illustrate the computational derivation of the response of some of the cohort statistics voice commands.





The second snapshot is an example of answering the question:

What is the total calorie intake of Florida males with BMI ≥ 28?

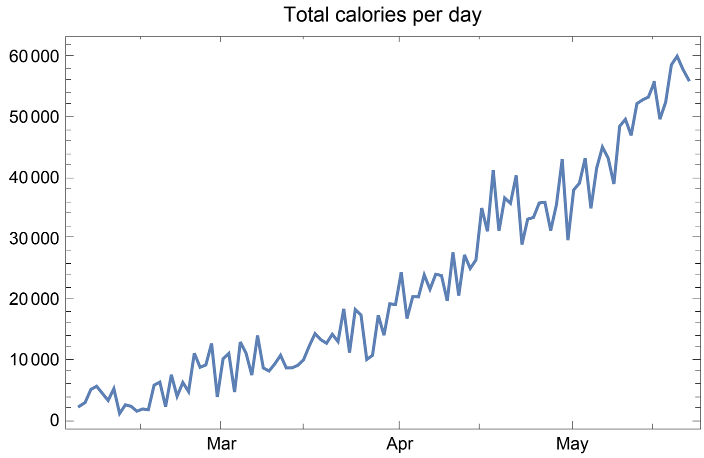
This computation is going to happen in the cloud and:
-- the time series graph can be displayed with Amazon Echo Show, and/or
-- a computation summary can be told through the standard Amazon Echo.

Code

Here is the actual Mathematica code:

In[195]:=
```
DynamicModule[
 {age = "all", bmi = 28, nel = "calories", state = "Florida", sex = "male"},
 (DateListPlot[#1, PlotLabel → "Total " <> nel <> " per day"] &) [
  TimeSeries[Values /@ Normal[
     tud[Select[(#state == state || state == "all") && (#sex == sex || sex == "both") &&
          (#age == age || age == "all") && #BMI ≥ bmi &]][GroupBy[#date &] /* Values,
      Association["date" → First /* "date", "total" → Query[Total, nel]]]]]]]]
```

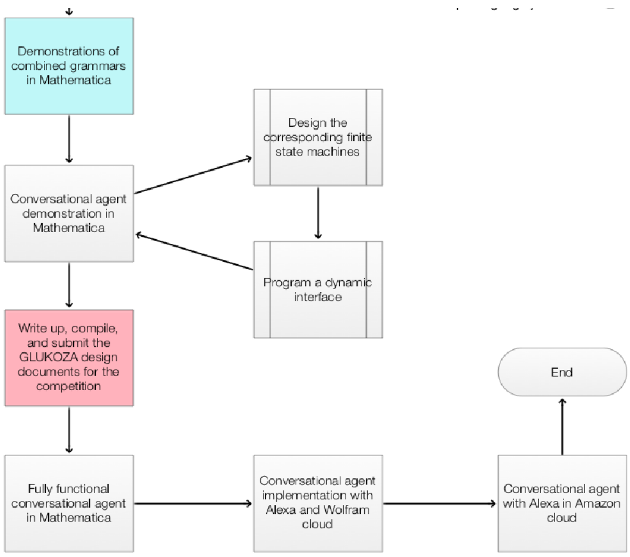
Out[195]=



Future plans

The overall development plan for GLUKOZA is shown in the submission document (diagram) ""GLUKOZA development stages",[2].

Here is the relevant diagram excerpt for next stages:



Most notably, after making a fully functional prototype in Wolfram Language (Mathematica) as shown in [9,10] we are going to proceed with making a full blown Alexa skill using Wolfram Cloud. (See [11].)

References

Additional submission documents

The following files are referenced in this document and are part of competition submission.

[1] “Morphological analysis of GLUKOZA’s conversational agent elements over functionalities-vs-data breakdown”.
File name “Morphological analysis of GLUKOZA’s CA elements over Functionalities-vs-Data.pdf” .

[2] “GLUKOZA development stages”, diagram.
File name: “GLUKOZA development stages.pdf”

[3] “GLUKOZA design”, diagram.
File name: “GLUKOZA design.pdf”

[4] “GLUKOZA conversational agent modular design”, mind-map.
File name : “GLUKOZA conversational agent modular design.pdf”

[5] “GLUKOZA grammars”, PDF of a Mathematica notebook.
File name: “GLUKOZA grammars.pdf”

Conversational agent design

[6] Anton Antonov, “Creating and programming domain specific languages”, (2016), MathematicaForPrediction at WordPress blog.
URL: https://mathematicaforprediction.wordpress.com/2016/03/22/creating-and-programming-dsls .

[7] Anton Antonov, Functional parsers, Mathematica package, MathematicaForPrediction at GitHub, 2014.
URL: https://github.com/antononcube/MathematicaForPrediction/blob/master/FunctionalParsers.m .

[8] Anton Antonov, "Natural language processing with functional parsers", (2014), MathematicaForPrediction at WordPress blog.
URL: https://mathematicaforprediction.wordpress.com/2014/02/13/natural-language-processing-with-functional-parsers .

[9] Anton Antonov, “Simple time series conversational engine”, (2014), MathematicaForPrediction at WordPress blog.
URL: https://mathematicaforprediction.wordpress.com/2014/11/29/simple-time-series-conversational-engine .

[10] Anton Antonov, “Phone dialing dialogs conversational agent”, (2017), ConversationalAgents at GitHub,
URL: https://github.com/antononcube/ConversationalAgents/tree/master/Projects/PhoneDialingDialogsAgent .

[11] Todd Gayley, Commanding the Wolfram Cloud, Wolfram Technology Conference 2015). URL: https://www.youtube.com/watch?v=kSTlY6fHCMk .
This 45 min movie shows how Echo and Alexa can be hooked up with the Wolfram Research Inc. technologies.

Diabetic users simulations

[12] United States Census Bureau, American Community Survey 5-Year Estimates: B01001, Sex by Age, American FactFinder. (2012).
URL: https://factfinder.census.gov/faces/tableservices/jsf/pages/productview.xhtml?src=bkmk .

[13] Centers for Disease Control and Prevention, Diabetes Atlas.

URL: https://gis.cdc.gov/grasp/diabetes/DiabetesAtlas.html .

[14] AACE/ACE COMPREHENSIVE DIABETES MANAGEMENT ALGORITHM, (2017).
URL: https://www.aace.com/publications/algorithm .

[15] Wikipedia entry: https://en.wikipedia.org/wiki/Diabetes_management .