BACKGROUND

I have started this project for several reasons, those are

1. To get some experience on Data scraping from online
2. To get some experience on Python “BeautifulSoup” package.
3. How to make own database and parse them
4. How to use tkinter table widget.
5. And overall to strengthen my python skill

IDB DATABASE

It is a custom made database system (for sake of trying). It is mainly python multidimensional list which is pickled by python pickle package. The first element of the list is a Dictionary. The list look like this [{},[],[]]. The first element of the list, which is a dictionary contain the title/header name and datatype. Currently it supports two types of data 1. String 2. Float. Integers are automatically converted to a floating point.

Example of idb data.

[{“id”:””,”name”:””,”salary”:”0”}, [“01”,”abcd”,1200.00],[“02”,”efgh”,1100.00]]

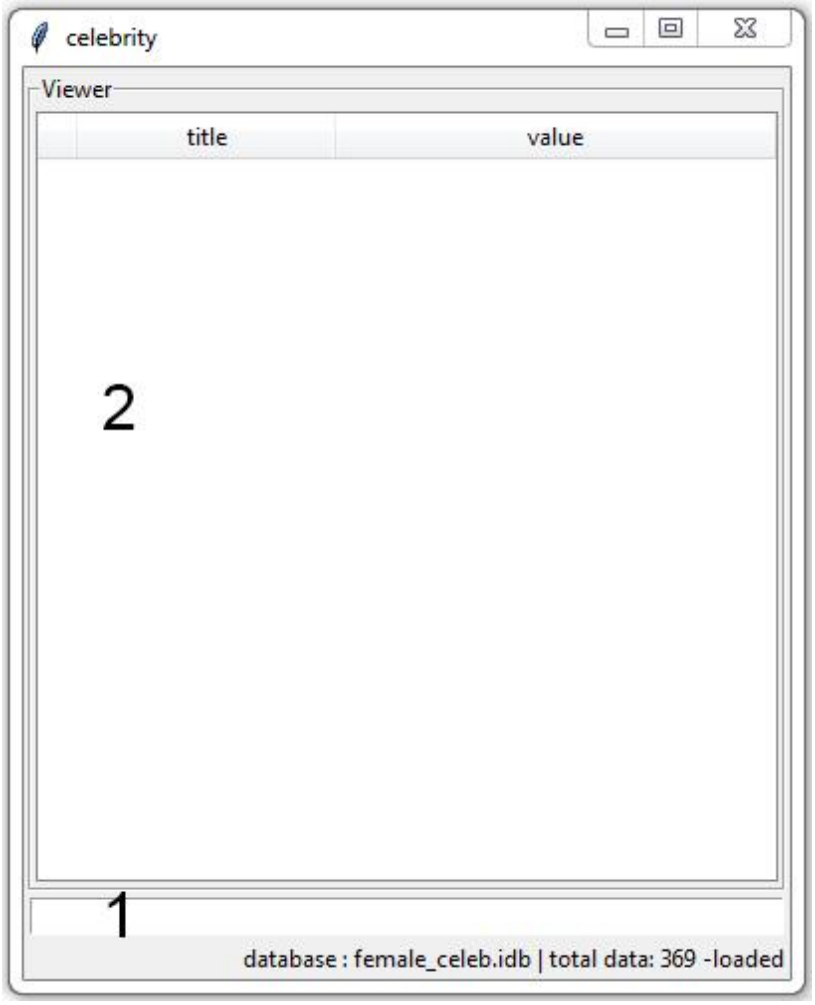
To explore the database use pickle.load()

LIMITATION

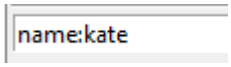
The “lookup.py” is the main module which handle this database, unfortunately this “lookup.py” module and to be exact the whole program is heavily customized to open “female_celeb.idb” data. In future I will try to add a General purpose idb handler module.

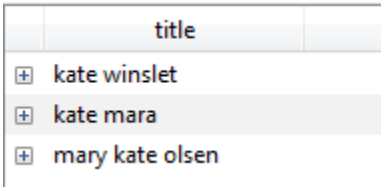
HOW TO USE THE PROGRAM

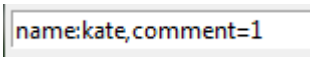
User can input mainly four Commands in input entry box (1 in image) they are Name, Input, Type, Fuzzy. After typing input one should must use “:” colon sign without any space. Then specific arguments next to that colon. The output of that command will be shown in table (2 in image). All the commands and respected arguments are described below.



Name

Command should look like this  first the command key “name” then colon “:” and the last is the argument of that command, here we use “kate” as an argument. Press enter to activate the command. It will return celebrity information which contain “kate” in their name.



You can use  “comment=1” and “comment=2” argument at the end of main name argument. Which will return estimated body type of that celebrity. Comment=1 is for short information, comment=2 is detail version. The body type ratios are in “classifier.py” module, you can change the ratio as your preferences.

Input

Command should look like this input:5.4 or input:5.4,34,27,36, here first one is only height, and second one is, height, chest, waist, and hip. If you use second way you must need to specify all three arguments with height value.

Let say someone's measurement is 5.4,35,27,36, this command will show which celebrities are sharing almost same body shape. Or how much modification needed to be a perfect body shape, or other five type shapes classified in classifier.py module.

Though height value looks like decimal, but it is not. Here 5.4 means 5 feet 4 inches. This system is wired, need to be changed. [In future]

You can also use centimeter value instead of inch and feet. There is a bug in conversation of unit. Need to fix.

At the end of argument, using "comment=1" or "comment=2" will provide comments and "thres=1" or any numeric value after "thres=" will give much for precise result. Default threshold is 5.

Type

Command should look like this type:perfect or type:big hip, there are 5 body type included in "classifier.py" module. You can add more option by including the name and ratio in that module. "comment=1 or 2" can be use here. you can change the ratio as your preferences.

Fuzzy- most useful

To make this program general purpose (work with any database) this method need to polish more.

It is actually multiple point search (not strictly fuzzy) and the most powerful command. You can use any of the command described before, or combined two or more commands. You can use data title/header as well. The listed title in "female_celeb.idb" are, height, weight, bra size, chest, waist, hip, birthday, nationality, occupation and eye.

title	value
kate winslet	
height	5 feet 6 inch
weight	63.05 kg
bra size	34C
chest	37.01 in
waist	26.77 in
hip	35.83 in
birthday	10/05/1975
nationality	UK
occupation	Actress
eye	blue

If you want to find celebrity with brown eyes, height 5 feet 5 and waist 27 inch then the fuzzy search will look like this `fuzzy:eye=brown,height=5.5,waist=27`

If you want more precise result you can use threshold, so it will look like this.

`fuzzy:eye=brown,height=5.5,waist=27,thres=1` Or with comment

`fuzzy:eye=brown,height=5.5,waist=27,thres=1,comment=1`

Extra

show:1 command will enable image viewer, and show:0 will disable the viewer

size:500,500 (or any x,y value) will change the viewer size.

save: command will save “show” and “size” value in option.opt

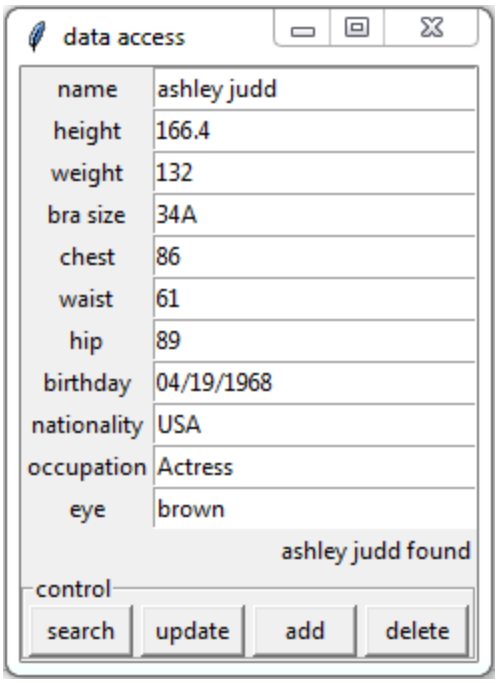
Python imaging library (pillow, pip install pillow) will required. For this commands.

Image enabled database required for this, the image title will contain the path of the image file. Image enabled database look like this, “image title contain the path of the image”.

title	value
Marlin	
height	5 feet 2 inch
weight	54.43 kg
bra size	34C (75C)
build	Natural
chest	34.0 in
waist	22.0 in
hip	32.0 in
body type	Slim
birthday	10/04/1997
nationality	American
ethnicity	Caucasian
hair	BrownLong, Straight
eye	Green
also known as	marry, maria
image	/image_dir/marlinG.jpg

“data_access.pyw” is a simple program to modify “female_celeb.idb” database and should work with other “idb” database where first element is the dictionary with title and type. (Maybe won’t work right now, bug need to be fixed)

In name entry box type the name and press search, it will return all the value the name contain. The name must need to be exact as stored in database for safety purpose. I



You can add new info by providing all the information.