# **Progress of the Internship – Isabel Soares**

12/7 - 20/7: Searching for tools that helps me to create a general chat bot.

[Presentation with many possibilities in attachment]

20/7 - 28/7: Developing a chat bot that answers a certain type of questions through the Python Library, ChatterBot.

#### **CHATTERBOT DOCUMENTATION:**

To install, follow these steps: <u>Installation — ChatterBot 1.0.8 documentation</u>

To start, I found this tutorial that helped me very much in understanding better how ChatterBot works: How to Create a Chatbot Using Chatterbot Python - YouTube

## **Limitations of ChatterBot:**

- a simple FAQs; (I found this limitation while implementing)
- we cannot filter the information through the question; (*I found this limitation while implementing*)
- we cannot have more than 1 answer per question. (*I found this limitation while implementing*)

## **Advantages of ChatterBot:**

- Easy to use: chatbot answers questions similar to the Q&A.json(that was created by me) (gunthercox/ChatterBot: ChatterBot is a machine learning, conversational dialog engine for creating chat bots (github.com) && ChatterBot Tutorial ChatterBot 1.0.8 documentation ) and easy to add the data to the bot. That's why this type of chat bot is good for FAQs, for instance.
- Have Corpus for many languages (<u>gunthercox/chatterbot-corpus: A multilingual dialog corpus (github.com)</u>).
- Use Search and Classifications Algorithms (<u>Frequently Asked Questions ChatterBot</u>
   1.0.8 documentation) which helps in compensating for users' typos:

#### Basically, chat bot selects a response include:

### [Search Algorithm]

- o the similarity of an input statement to known statements
- o the frequency in which similar known responses occur
- the likeliness of an input statement to fit into a category that known statements are a part of.

#### [Classification Algorithm]

 Use naive Bayesian classification algorithms to determine if an input statement meets a particular set of criteria that warrant a response to be generated from that logic adapter.

## Improvements on my implementation of the ChatterBot:

- Not a random insight, but a string concatenated.

#### SIMPLE CONVERSATION EXAMPLE

Figure 1 - A conversation using Chatterbot.

**28/7 – Present:** Developing a chat bot that answers a certain type of questions through an open source tool, called Rasa.

#### **RASA DOCUMENTATION:**

Download: Installing Rasa Open Source: Windows 10 (Updated 2021) - YouTube

Quick Installation - <u>Installation (rasa.com)</u>

Note: I needed to use administrator permissions, because Anaconda did not work without it.

I started to understand the documentation but, in my opinion, it is more useful to watch these videos from the Rasa YouTube Channel, although they show an old version (the main changes between the versions are in the format used, which was previously on Markdown and now is on YAML and the last version is easier and simpler than first, regarding to the policies and pipeline): <a href="https://www.youtube.com/watch?v=-">https://www.youtube.com/watch?v=-</a>

F6h43DRpcU&list=RDCMUCJ0V6493mLvqdiVwOKWBODQ&start radio=1&t=12s

In summary, Rasa divides itself in three majors' components:

- 1. <u>Rasa NLU</u> It is responsible for understanding the input messages of the user. It should recognize what the user wants and also extract information like names, places from the messages.
  - a. Domain: defines the universe in which your bot operates. It specifies exactly:
    - i. which intents you are expecting to respond to
    - ii. which slots you wish to track
    - iii. which actions your bot can take
  - b. Actions: are the things your bot can do.
  - c. Stories: is a training data sample for the dialogue system. This shows the bot how to act and react to the inputs given by the user.
  - d. Slots: are the things you want to keep track of during a conversation. The slots are the bot's memory.
- 2. Rasa Core It maintains the conversation flow between the User and the Bot.
- 3. Rasa X is a toolset for developers to build, improve, and deploy contextual assistants with the Rasa framework. The good thing about Rasa X is that you can share your assistant with real users and collect the conversations they have with the bot, allowing you to improve it without interrupting the bot running in production.

Good summarize in: <u>How to build a contextual assistant using RASA Forms | by Amal Menzli | Towards Data Science</u> (I found very useful to understand, in a general way, the documentation of Rasa)

When you are running/training your bot use an Anaconda Powershell Prompt and the **list of commands** that you can use are listed in <u>Command Line Interface (rasa.com)</u>

A good way to understand how querying a database with a chatbot works is by watching this livecoding: Rasa Livecoding: Querying a database with a chatbot - YouTube and the code is here (also very useful to understand how Rasa works with this example): GitHub - rctatman/personal website bot: A simple bot to answer questions on my personal website. (In development)

## **Limitations/Disadvantages of Rasa:**

- more difficult in terms of understanding all the documentation; (<u>Introduction to Rasa Open Source</u>)
- more difficult in terms of understanding how to connect our data to the bot, because it creates a database from our data (in my case .csv) (*I found this limitation while implementing and the links above helped me to overcome this limitation*)
- this is a limitation of rasa but I do not know the reason. I will give a tip that always helped me: when something is not working and you do not understand the bug, the best solution is to delete the models and the database of the obligations and insights. After, run again *rasa train* (to have a new a model) and create again new databases running the Jupiter notebook.

## **Advantages of Rasa:**

- more powerful and complex tool than chatterbot, because the code of rasa keeps being updated and chatterbot is no longer updated (*I found this limitation comparing the documentation of both tools*);
- it also has a similar algorithm to ChatterBot, which tries to compensate for users' typos.

I started by creating a simple conversation, like a FAQs where the bot only answers two questions:

Figure 2 - A simple conversation using Rasa.

By exploring some more functionalities of Rasa, I concluded that sometimes if the keyword was wrongly written, the bot will not understand and you can see it here:

```
Account Promoted Name Management (1985) | College | Coll
```

Figure 3 - A limitation of Rasa.

(NOTE: At the left Anaconda Powershell Prompt is where the actions associated to the database run and I added some prints to test what and if the bot understood the keyword even if it was not correctly written and at the right, you can see the conversation between bot and user)

<u>Both with Chatterbot and Rasa</u>, I used some **type questions** regarding the *obligations*, such as:

- How much is my VAT?
- How much I need to pay regarding VAT?
- How much is my Social Security?
- How much I need to pay regarding Social Security?
- How much is my Deductions?
- How much I need to pay regarding Deductions?
- How much is my Compensation Fund?
- How much I need to pay regarding Compensation Fund?
- Until when can I pay my VAT?
- What is the limit of date that I can pay my VAT?
- Until when can I pay my Social Security?
- What is the limit of date that I can pay my Social Security?
- Until when can I pay my Deductions?
- What is the limit of date that I can pay my Deductions?
- Until when can I pay my Compensation Fund?
- What is the limit of date that I can pay my Compensation Fund?

And related to the *insights*, I used these:

- Show me a positive insight.
- Show me a medium insight.
- Show me a negative insight.
- Show me all insights.
- Show me all positive insights.
- Show me all medium insights.
- Show me all negative insights.

After querying a database with a chatbot, I was able of creating a more complex conversation, where the bot answers multiple type questions (some displayed above):

```
(Case) PS ("Utern'12seb. Sepres/Desire) (herebothPRI2.0) ress shell

8221-88-13 10-18:08 PMO

8221-88-13 10-18:08 PMO

8221-88-13 10-18:08 PMO

8221-88-13 10-18:08 PMO

1 given channels, ont the

1-connector' argument. Any other channels will be ignored. To connect to all

1 given channels, ont the

1-connector' argument.

1 given channels, ont the

1 given channels, on
```

Figure 4 - More complex conversation using Rasa.

More complete conversation with some limitations on date key (only works with these expressions "April 2021" and "May 2021"):

```
Description of the property of
```

Figure 5 - More complete conversation with some limitations yet (using rasa).

Trying to ask the bot the payment date for each obligation, the bot wrongly answered the value to pay. I supposed that it could be a training problem (maybe these two actions are very similar to the bot), in order to understand how to solve this bug, I used a rasa command, rasa interactive, which helped me to see the confidence of each action and each intent. (The prints show the second time that I run it with this command)

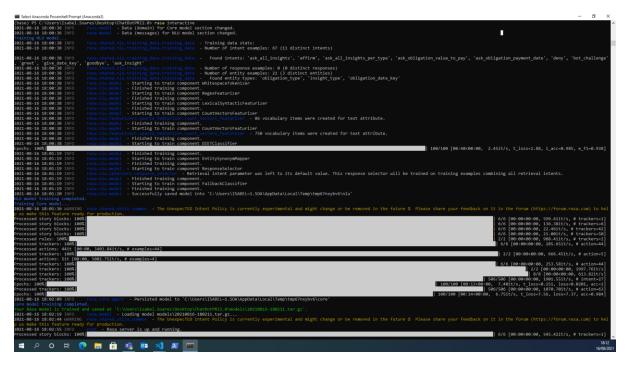


Figure 6 - Start running the command rasa interactive.

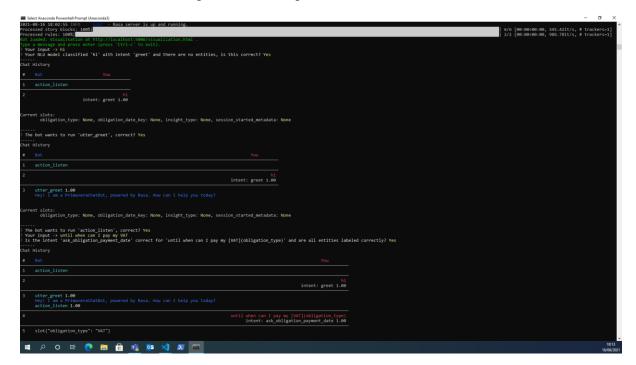


Figure 7 - Running rasa interactive (continuation).

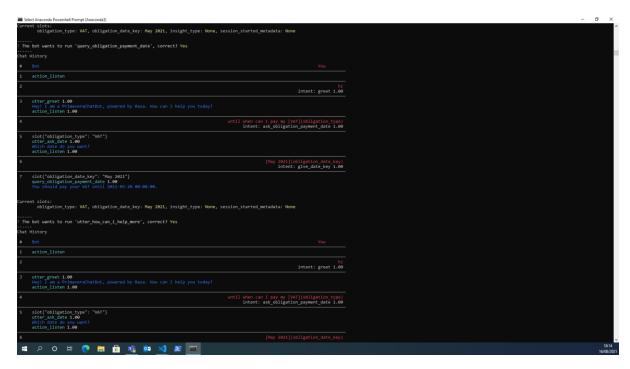


Figure 8 - Running rasa interactive (continuation).

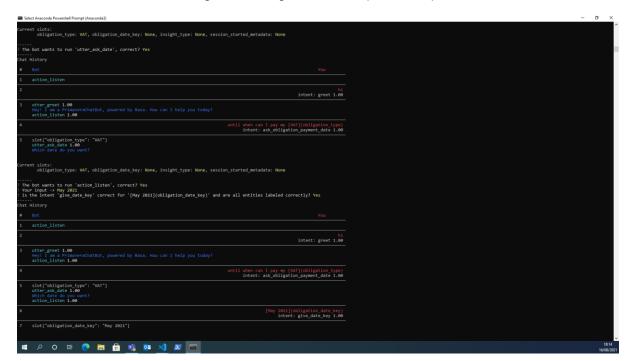


Figure 9 - Running rasa interactive (continuation).

Note that the value that appears next to the actions and intents is the confidence that the bot has in each of them.

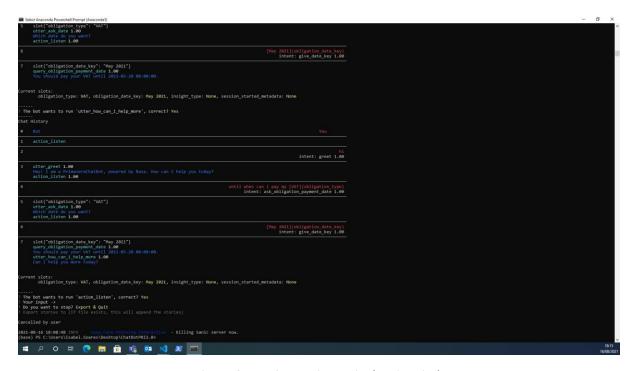


Figure 10 - Running rasa interactive (continuation).

(Before these prints) 1 noticed that the confidence of the action "query\_obligation\_payment\_date" was 0.22 and that the confidence of the action "query\_obligation\_value\_to\_pay" was 0.30. This is the reason why the bot always answered the value to pay and not the payment date. In order to solve this bug, I gave more confidence to the action "query\_obligation\_payment\_date". However, after the bot started always answering the payment date and not the value to pay.

While running *rasa interactive*, I could see a visualization of what was happening in real time with the bot:

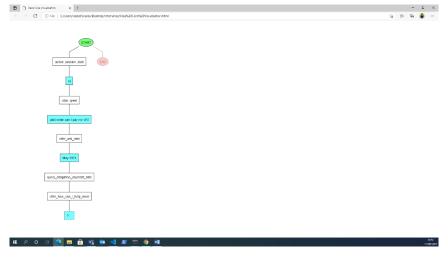


Figure 11 - Rasa visualization of the story.

Note: If you run the command *rasa visualize,* it generates a visual representation of your stories, like the one above.

After many trainings using *rasa interactive* and after changing the names of the actions and intents (because they were similar and this confused the bot), the bot was able to answer all the questions correctly:

```
Association Processing Continues and Continu
```

Figure 12 - The bot answers questions about the obligations.

```
The property of the property (accounts)

For 19 to 10 control (10 control (10
```

Figure 13 - The bot answers questions about the insights.

Then, in order to solve the limitation on date key (the user had to write in a specific way for the bot to understand), I used a Python library, *dateutil*. Hence, the bot is able of parsing different date representations. (You can see on the picture below):

```
A manuford homest from the control of the control o
```

Figure 14 - A conversation with different date formats and issues on certain keywords.

Still regarding the picture above, I could notice that when I/ the user write/s the keyword wrongly, per default, the bot answers as if a positive insight. Maybe because this value is the first on the categorical slot, <code>insight\_type</code>.