Getting Started Documentation

This document is intended to accompany the project's main dissertation *DataBot: A Conversation System for Sourcing Data in CKAN,* Rasa Core, Rasa NLU, CKAN and Supervisor documentation. **PLEASE READ SUPPORTING DOCUMENTATION BEFORE STARTING.**

Before development, it is **recommended** that you complete the CKAN Extension tutorial and Rasa Core tutorials.

Overview:

1. Deploy
    A. Installing the CKAN Extension
    B. Installing Rasa
    C. Install and run the Rasa Agent
    D. Notes
2. Develop
    A. Train Dialogue
    B. Train NLU
    C. Test & Evaluation
    D. Notes
3. Useful links

1. Deploy

A. Installing the CKAN Extension

1. If you haven't Installed a CKAN instance on your machine, install one by following the CKAN Installation guide.
2. If you haven't deployed the CKAN instance before, deploy your instance by following Deploying a source install..
3. Clone branch `data-bot-ckan` from Landscape4Data Github repository.
4. In your CKAN config file, search for `ckan.plugins` and append `rasa` at the end of the line, separated by a space.
5. Run command `sudo service apache2 restart`. If that doesn't work, try `sudo /etc/init.d/apache2 restart`.

B. Installing Rasa

1. In `src/ckanext-rasa`, run command `pip install -r requirements.txt`. Alternatively, you could use commands `pip install rasa-core` and `pip install rasa-nlu[spacy]`. Read Rasa Core Installation Guide for more information. This will install both Rasa Core and Rasa NLU with the spaCy extension.
2. Once rasa-nlu is installed, run the following commands in your terminal; `python -m spacy download en_core_web_md` and `python -m spacy link en_core_web_md en`. This will download the word embeddings and create a link to it.

C. Install and run Rasa Core Server

1. If you haven't installed `supervisor`, install it by running the following commands in your terminal; `sudo apt-get install supervisor` and `service supervisor restart`.
2. Next, run command `sudo nano /usr/local/bin/rasa_agent.sh` and insert the following code:

```
#!/bin/bash
python <your ckan source directory>/ckanext-rasa/ckanext/rasa/server.py
```

   You can run `python <your ckan source directory>/ckanext-rasa/ckanext/rasa/server.py -h` for more information on arguments that can be supplied.
3. Run command `chmod +x /usr/local/bin/rasa_agent.sh`
4. Run `sudo nano /etc/supervisor/conf.d/rasa_agent.conf` where `/etc/supervisor/conf.d` is your default supervisor directory and insert the following code:

```
[program:rasa_agent]
command=/usr/local/bin/rasa_agent.sh
autostart=true
autorestart=true
stderr_logfile=/var/log/rasa_agent.err.log
stdout_logfile=/var/log/rasa_agent.out.log
```

   Useful link might be found here.

5. Run command `sudo supervisorctl reread` followed by `sudo supervisorctl update`

D. Note

1. After completing **ALL** the 3 steps above, you should be able to browse *http://udlest1.cs.ucl.ac.uk/databot* or <your *hostname>/databot.* **Due to firewall restrictions,** data sourcing will not work unless you are on UCL's network.

2. Develop

This section will assume you have completed the aforementioned tutorials. If you would like to find out more about the code, please read the written dissertation Chapters 4 and 5. Additionally, have a look at the code yourself!

A. Train Dialogue

1. Create Stories by following the Story format defined here.
2. Save the Stories in at `<your source directory>/ckanext-rasa/ckanext/rasa/data_bot/main/data/stories.md`
3. Run command

```
python <your source directory>/ckanext-rasa/ckanext/rasa/data_bot/main/bot/develop.py dialogue
```

B. Train NLU

1. Create NLU Training data following the format defined here.
2. Save your data at `<your source directory>/ckanext-rasa/ckanext/rasa/data_bot/main/data/enchanced-data.json`
3. Run command

```
python <your source directory>/ckanext-rasa/ckanext/rasa/data_bot/main/bot/develop.py nlu
```

C. Test & Evaluate

1. To test dialogue, run command

```
python -m rasa_core.evaluate -d  <your source directory>/ckanext-rasa/ckanext/rasa/data_bot/main/models/dialogue -s  <your source directory>/ckanext-rasa/ckanext/rasa/data_bot/main/data/<your test stories>
```

Use `-h` for more info on `rasa_core.evaluate`.

2. To test NLU, run command

```
python -m rasa_nlu.evaluate -d directory>/ckanext-rasa/ckanext/rasa/data_bot/main/data/<your test nlu data> -m <your source directory>/ckanext-rasa/ckanext/rasa/data_bot/main/models/<your nlu model>
```

Use `-h` for more info on `rasa_nlu.evaluate`. Look at Rasa NLU Evaluation docs for more detail.

3. To run unit tests, run command

```
python -m unittest discover  <your source directory>/ckanext-rasa/ckanext/rasa/
```

D. Notes

1. If your system does not have a graphical interface ie. a server, then you would need to edit rasa_core.evaluate and rasa_nlu.evaluate to obtain the confusion matrices, precision, recall and f1-score. Look at this Stack Overflow post for more info. Also, the rasa_nlu.evaluate by default does not export the confusion matrix - edit the module at the necessary line with this tip.

3. Useful links
    a. Automated Data Formatters
        i. Tracy
        ii. Chatito
    b. How to Install CKAN 2.7.2 on Vagrant (good workflow setup for windows)
    c. Rasa Core Github Issues and Gitter pages