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](https://core.rasa.com/), [Rasa NLU](https://nlu.rasa.com/), [CKAN](http://docs.ckan.org/en/ckan-2.7.3/) and [Superviso](http://supervisord.org/)r documentation. **PLEASE READ SUPPORTING DOCUMENTATION BEFORE STARTING.**

Before development, it is **recommended** that you complete the [CKAN Extension tutorial](http://docs.ckan.org/en/latest/extensions/tutorial.html#) and [Rasa Core tutorials.](https://core.rasa.com/tutorial_basics.html)

Overview:

1. Deploy
2. Installing the CKAN Extension
3. Installing Rasa
4. Install and run the Rasa Agent
5. Install and run Redis
6. Notes
7. Develop
   1. Train Dialogue
   2. Train NLU
   3. Test & Evaluation
   4. Notes
8. 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](http://docs.ckan.org/en/ckan-2.7.3/maintaining/installing/install-from-source.html) guide.
2. If you haven’t deployed the CKAN instance before, deploy your instance by following [Deploying a source install](http://docs.ckan.org/en/latest/maintaining/installing/deployment.html#create-the-wsgi-script-file)..
3. Clone branch data-bot-ckan from [Landscape4Data](https://github.com/mgalas/Landscape4Data) Github repository.
4. In your CKAN config file, search for ckan.pluginsand 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](https://core.rasa.com/installation.html) for more information. This will install both Rasa Core and Rasa NLU with the [spaCy](https://spacy.io/) 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.

1. Run command chmod +x /usr/local/bin/rasa\_agent.sh
2. 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](https://www.digitalocean.com/community/tutorials/how-to-install-and-manage-supervisor-on-ubuntu-and-debian-vps).

1. Run command sudo supervisorctl reread followed by sudo supervisorctl update

D. Install and run Redis Server

1. Redis comes installed with CKAN.
2. Run Redis with command sudo redis-server --daemonize yes. For a more robust method, you can create a configuration file instead. Read more [here](https://redis.io/topics/quickstart).

E. Note

1. After completing **ALL** the 3 steps above, you should be able to browse [*http://udlest1.cs.ucl.ac.uk/databot*](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!

1. Train Dialogue
   1. Create Stories by following the Story format defined [here](https://core.rasa.com/stories.html).
   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

sudo supervisorctl

restart rasa\_agent

1. Train NLU
   1. Create NLU Training data following the format defined [here.](https://nlu.rasa.com/dataformat.html)
   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

sudo supervisorctl

restart rasa\_agent

1. 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.

1. 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](https://nlu.rasa.com/evaluation.html) docs for more detail.

1. To run unit tests, run command

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

1. Notes
2. 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](https://stackoverflow.com/questions/2801882/generating-a-png-with-matplotlib-when-display-is-undefined) 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](https://stackoverflow.com/questions/9012487/matplotlib-pyplot-savefig-outputs-blank-image) tip.

3. Useful links

* 1. Automated Data Formatters
     1. [Tracy](https://yuukanoo.github.io/tracy/#/agents)
     2. [Chatito](https://rodrigopivi.github.io/Chatito/)

# [How to Install CKAN 2.7.2 on Vagrant](https://github.com/ckan/ckan/wiki/How-to-Install-CKAN-2.7.2-on-Vagrant) (good workflow setup for windows)

* 1. Rasa Core [Github Issues](https://github.com/RasaHQ/rasa_core/issues) and [Gitter](https://gitter.im/RasaHQ/rasa_core) pages
  2. [Visualize stories](https://core.rasa.com/stories.html#visualization-of-stories)