

## Installation instructions:

Download and install Docker (Community Edition / Stable channel) following the instructions on:

<https://docs.docker.com/engine/installation/>

After Docker is running on your machine, run the following command to get the Docker container for this tutorial on your machine:

```
docker pull melcutz/nlu-demo
```

Because of all the dependencies (Spark, SpaCy, NLTK, UMLS, etc.) the image file is very big (13.4GB) so it might take a while to download.

Once the command finishes successfully and you have the image on your machine (use 'docker images' to validate), use the following command to start the Docker container:

```
docker run -it --rm -p 8888:8888 melcutz/nlu-demo
```

If successfully launched, the output should be something like:

```
Copy/paste this URL into your browser when you connect for the first time,  
to login with a token:  
http://localhost:8888/?token=a8309a652c58fe0172483ef845461af030349e04cb0ac88e
```

so, follow the instruction and copy/paste the provided URL into your browser of choice (we tested on Chrome) and you should be able to navigate to an instance of Jupyter Notebook running inside the Docker container. The browser content should be like the following picture.

localhost:8888/tree

jupyter
Logout

FilesRunningClusters

Select items to perform actions on them.
UploadNew

	Name ↑	Last Modified ↑
<input type="checkbox"/>	demo-data	36 minutes ago
<input type="checkbox"/>	nlTK_data	35 minutes ago
<input type="checkbox"/>	pysparknlp-1.0.0	a day ago
<input type="checkbox"/>	QuickUMLS	38 minutes ago
<input type="checkbox"/>	simstring	37 minutes ago
<input type="checkbox"/>	Solutions	38 minutes ago
<input type="checkbox"/>	1_spacy_Intro.ipynb	4 days ago
<input type="checkbox"/>	2_spacy_pubmed.ipynb	4 days ago
<input type="checkbox"/>	3_spacy_pubmed_model.ipynb	4 days ago
<input type="checkbox"/>	4_SparkNLP_intro.ipynb	4 hours ago
<input type="checkbox"/>	5_SparkNLP_pubmed_model.ipynb	4 hours ago
<input type="checkbox"/>	architecture.svg	a month ago
<input type="checkbox"/>	df_matches_pubmed_diabetes_100.data	a month ago
<input type="checkbox"/>	df_matches_pubmed_diabetes_1000.data	a month ago
<input type="checkbox"/>	df_matches_pubmed_diabetes_5000.data	a month ago
<input type="checkbox"/>	disease_drugs.png	a month ago
<input type="checkbox"/>	helper_methods.py	4 days ago
<input type="checkbox"/>	negex.py	7 months ago
<input type="checkbox"/>	negex_triggers.txt	7 months ago
<input type="checkbox"/>	pipeline.svg	a month ago
<input type="checkbox"/>	pysparknlp-1.0.0.tar.gz	4 hours ago
<input type="checkbox"/>	SemGroups_2013.txt	7 months ago
<input type="checkbox"/>	setup_simstring.sh	a month ago
<input type="checkbox"/>	strata-requirements.txt	4 hours ago

Congratulations! You're all set for the tutorial!