

# **OBJECTIVE: CLASSIFICATION OF PATENT ON THE BASIS OF TOPIC OF PATENT**

## **Problem Statement:**

Given a set of patents filed on different topics train a classifier or use a tool which when provided with an user input of a patent can provide a judgement as to what topic the patent is on as well as the probabilities that the patent is talking on a set of topics.

I collected four sets of patent titles 'Cybersecurity', 'Data Mining', 'Solar Energy' and 'Graphene' from Google Patents. I then used an open source tool called RasaNLU to classify the patents based on the patent title.

Link to github repo: <https://github.com/shaliniit/Kreatrade-NLP-Patent-Corpus>

## **Steps:**

- 0.Open cmd.
- 1.Mkdir Desktop\PatentBot
- 2.cd PatentBot
- 3.Run npm i -g rasa-nlu-trainer in a second cmd window which has been opened in Administrator mode.
- 4.Run python -m spacy download en in a second cmd window which has been opened in Administrator mode.
- 5.mkdir data in first cmd window
- 6.cd data
- 7.echo "->data.json and the open data.json file under data.Copy paste the contents of the data file in git repository.(Cloning the git repository can save the time of copying all files actually).
- 8.rasa-nlu-trainer in first cmd window.
- 9.Enter the training data in the server window which pops up by clicking on Add Example at the top and then click on Save.
- 10.Open another third cmd window while the server runs on this window.
- 11.Create config\_spacy file and copy paste code from Git Repository.
- 12.Copy paste requirements.txt from Git Repository.
- 13.Copy paste nlu\_model.py file from Git Repository.
- 14.Run pip install -r requirements.txt in same cmd window.

15.Run nlu\_model.py to create a folder called models in PatentBot which will contain all pipeline related files.

16.Enter input in nlu\_model.py file in line 14 Interpreter.parse field.

17.Run nlu\_model.py again to see the results.

18.Enter every new test case in the server which is running.

19.Run train\_nlu as well as run\_nlu every time so that the model gets trained with every new test case.

20.Copy results to another file if needed.

### **Accuracy:**

92.8%

18 errors in 250 test cases with 123 training examples. Optimum would be around 450-500 training data cases. It gives almost no error.

### **Suggestions:**

1.Increasing the number of training examples.

2.Running both train\_nlu and run\_nlu each time. It takes more time but trains the model every time.

3.Adding every test case encountered to the training data in the server.

### **Useful Links:**

1. <https://rasa.com/docs/rasa/>
2. [https://forum.rasa.com/?\\_ga=2.219041177.1002252592.1562763324-389945078.1558386011](https://forum.rasa.com/?_ga=2.219041177.1002252592.1562763324-389945078.1558386011)
3. <https://github.com/shaliniit/Kreatrade-NLP-Patent-Corpus>

### **Images:**

FOLDERS

- ▼ react
- ▶ todoapp

```
1 {
2   "rasa_nlu_data": {
3     "regex_features": [
4       {
5         "name": "Solar Energy",
6         "pattern": "solar.*"
7       },
8       {
9         "name": "Cybersecurity",
10        "pattern": "cybersecurity.*"
11      },
12      {
13        "name": "Graphene",
14        "pattern": "graphene.*"
15      },
16      {
17        "name": "Data Mining",
18        "pattern": "data mining.*"
19      },
20      {
21        "name": "Solar Energy",
22        "pattern": "Solar.*"
23      },
24      {
25        "name": "Cybersecurity",
26        "pattern": "Cybersecurity.*"
27      },
28      {
29        "name": "Graphene",
30        "pattern": "Graphene.*"
31      },
32      {
33        "name": "Data Mining",
34        "pattern": "Data mining.*"
35      }
36    ],
37    "common_examples": [
38      {
39        "text": "Booster unit, power o
40        "intent": "ask for power"
```

PatentBot

File

Home

Share

View

< > ▾ ↑ > This PC > Desktop > PatentBot >

★ Quick access

Desktop

Downloads

Documents

Pictures

Bankbot

data

data

weathernlu

Dropbox

OneDrive

This PC

CD Drive (E:) MobileV

Network

Name

data

models

config\_spacy

nlu\_model

requirements

Date modified

02-07-2019 00:2



02-07-2019 09:3

02-07-2019 09:2
















08-07-2019 19:4

28-05-2019 10:0

5 items

To see favorites here, select  then , and drag to the Favorites Bar folder. Or imp

C:\Users\Shalini\Desktop\PatentBot\data\data.json

Intent  		Text  	
	Solar Energy		Booster unit, power conditioners, and s
	Solar Energy		Direct beam solar lighting system
	Solar Energy		Demounting of inverted metamorphic m
	Solar Energy		Thin-film solar reflectors deployable fro
	Solar Energy		Solar panel
	Solar Energy		Reactive power compensation in solar
	Graphene		3-dimensional graphene structure and
	Graphene		Resin plating method using graphene
	Graphene		Preparation method for three-dimensio
	Graphene		Doped graphene electronic material
	Graphene		A method graphene reinforced aluminu
	Graphene		Method for manufacturing graphene-co

C:\> Administrator: Command Prompt

```
    r = requests.get(url)
File "C:\Users\Shalini\AppData\Local\Programs\Python\Python37\lib\site-packages\requests\models.py", line 91, in request
    return request('get', url, params=params, **kwargs)
File "C:\Users\Shalini\AppData\Local\Programs\Python\Python37\lib\site-packages\requests\session.py", line 530, in request
    return session.request(method=method, url=url, **kwargs)
File "C:\Users\Shalini\AppData\Local\Programs\Python\Python37\lib\site-packages\requests\session.py", line 646, in request
    resp = self.send(prepare_request(method=method, url=url, **kwargs))
File "C:\Users\Shalini\AppData\Local\Programs\Python\Python37\lib\site-packages\requests\session.py", line 698, in send
    r = adapter.send(request, **kwargs)
File "C:\Users\Shalini\AppData\Local\Programs\Python\Python37\lib\site-packages\requests\adapters.py", line 470, in send
    raise ConnectionError(e, request=request)
requests.exceptions.ConnectionError: HTTPSConnectionPool: Max retries exceeded with url: https://api.openai.com/v1/completions (Caused by urllib3.exceptions.MaxRetryError: HTTPSConnectionPool: Max retries exceeded with url: https://api.openai.com/v1/completions (Caused by urllib3.exceptions.NewConnectionError: <urllib3.connection.VerifiedHTTPSConnection object at 0x0000019000000000>: Failed to establish a new connection: [Errno 1006] Connection reset by peer))
C:\WINDOWS\system32>python -m spacy download en
Requirement already satisfied: en_core_web_sm==2.1.0 from C:\Users\Shalini\AppData\Local\Programs\Python\Python37\lib\site-packages in C:\Users\Shalini\AppData\Local\Programs\Python\Python37\lib\site-packages
[+] Download and installation successful
You can now load the model via spacy.load('en_core_web_sm')
symbolic link created for C:\Users\Shalini\AppData\Local\Programs\Python\Python37\lib\site-packages\spacy\en_core_web_sm
[+] Linking successful
C:\Users\Shalini\AppData\Local\Programs\Python\Python37\lib\site-packages>python -m spacy download en
-->
C:\Users\Shalini\AppData\Local\Programs\Python\Python37\lib\site-packages>python -m spacy download en
You can now load the model via spacy.load('en')

C:\WINDOWS\system32>npm i -g rasa-nlu-trainer
npm WARN registry Using stale data from https://registry.npmjs.org/ during revalidation.
C:\Users\Shalini\AppData\Roaming\npm\rasa-nlu-trainer -> npm WARN slick-carousel@1.8.1 requires a peer of jquery@1.9.0 but none is installed. You may need to run `npm install jquery` before installing this package.
npm WARN optional SKIPPING OPTIONAL DEPENDENCY: fsevents@1.2.9: SKIPPING OPTIONAL DEPENDENCY: Unsupported platform for fsevents@1.2.9
+ rasa-nlu-trainer@0.2.7
updated 2 packages in 156.159s

C:\WINDOWS\system32>python -m spacy download en
```

