


```
[308, 314, 'Companies worked at'],
[234, 240, 'Companies worked at'],
[175, 198, 'Companies worked at'],
[93, 136, 'Email Address'],
[39, 48, 'Location'],
[13, 37, 'Designation'],
[0, 12, 'Name']]]}]
```

```
len(cv_data)
```

```
200
```

```
!python -m spacy init fill-config /content/Resume-Parser/data/training/base_config.cfg /content/Resume-Parser/data/training/config.cfg
```

```
2023-09-30 07:32:20.997956: W tensorflow/compiler/tf2tensorrt/utlis/py_utils.cc:38] TF-TRT Warning: Could not find TensorRT
✓ Auto-filled config with all values
✓ Saved config
/content/Resume-Parser/data/training/config.cfg
You can now add your data and train your pipeline:
python -m spacy train config.cfg --paths.train ./train.spacy --paths.dev ./dev.spacy
```

```
cv_data
```

deployment of available resources to achieve organizational objectives • Mentoring the task force and ensuring development of team for service excellence IT Skills - Conversant with MS Office - Word, PowerPoint, Access and Excel. - Well versed with Internet Applications",

```
{'entities': [[9242, 9311, 'Skills'],
[9156, 9160, 'Graduation Year'],
[9109, 9128, 'College Name'],
[9070, 9107, 'Degree'],
[9064, 9068, 'Years of Experience'],
[9004, 9025, 'College Name'],
[8947, 9002, 'Degree'],
[6450, 6454, 'Years of Experience'],
[6418, 6427, 'Location'],
[6399, 6414, 'Companies worked at'],
[5900, 5909, 'Location'],
[5881, 5896, 'Companies worked at'],
[5472, 5481, 'Location'],
[5453, 5468, 'Companies worked at'],
[4811, 4820, 'Location'],
[4792, 4807, 'Companies worked at'],
[4127, 4136, 'Location'],
[4108, 4123, 'Companies worked at'],
[3939, 3942, 'Designation'],
[3825, 3834, 'Location'],
[3806, 3821, 'Companies worked at'],
[3770, 3773, 'Designation'],
[2236, 2245, 'Location'],
[2217, 2232, 'Companies worked at'],
[2191, 2194, 'Designation'],
[598, 613, 'Companies worked at'],
[593, 596, 'Designation'],
[100, 152, 'Email Address'],
[57, 66, 'Location'],
[52, 55, 'Designation'],
[22, 55, 'Designation'],
[0, 21, 'Name']]]}]
```

['Pradeep Kumar Security Analyst in Infosys - Career Contour Hyderabad, Telangana, Telangana - Email me on Indeed: [indeed.com/r/Pradeep-Kumar/96485546eadd9488](https://www.indeed.com/r/Pradeep-Kumar/96485546eadd9488) WORK EXPERIENCE Security Analyst in Infosys Career Contour - March 2015 to Present Security incident and event management (SIEM) is the process of identifying, monitoring and analysing security events or incidents within a real-time IT environment. It provides centralized view of the security scenario of an IT infrastructure EDUCATION B- Tech Jawaharlal Nehru Technological University - Hyderabad, Telangana SKILLS Splunk, Network Security, Arc sight (2 years), SIEM (2 years), McAfee ESM (1 year) ADDITIONAL INFORMATION Technologies: McAfee SIEM, Arc sight, HP Service manager. Description: Security incident and event management (SIEM) is the process of identifying, monitoring and analysing security events or incidents within a real-time IT environment. It provides centralized view of the security scenario of an IT infrastructure. Roles and Responsibilities: • Have Experience in device integration, alarm creation in ESM. • Having Experience on analysing FIREWALL events. • Checking health status for all devices in ESM. • Working in Offshore SOC team, Monitoring of SOC events, detecting and preventing the Intrusion attempts. • Hands-on experience monitoring the SIEM tool. • Provide level 2 supports to manage SIEM components. • Creating reports as per client requirement. • Analysing daily, weekly and monthly reports. • Creating case for the suspicious issue and forwarding it to Onsite SOC team for further investigation. • Monitoring the events on SIEM tool. • Creating the tickets in ticketing tool. <https://www.indeed.com/r/Pradeep-Kumar/96485546eadd9488?isid=rex-download&ikw=download-top&co=IN> <https://www.indeed.com/r/Pradeep-Kumar/96485546eadd9488?isid=rex-download&ikw=download-top&co=IN> • Finding false positive, fine tuning and escalating Security events.',

```
{'entities': [[587, 669, 'Skills'],
[512, 553, 'College Name'],
[503, 510, 'Degree'],
[177, 193, 'Designation'],
[34, 58, 'Companies worked at'],
[14, 30, 'Designation'],
[0, 13, 'Name']]]}]
```

```
def get_spacy_doc(file,data):
    np=spacy.blank('en')
    db=DocBin()
```

```

for text,annot in tqdm(data):
    doc=nlp.make_doc(text)
    annot=annot['entities']

    ents=[]
    entity_indices=[]

    for start,end,label in annot:
        skip_entity=False
        for idx in range(start, end):
            if idx in entity_indices:
                skip_entity=True
                break
        if skip_entity==True:
            continue
        entity_indices=entity_indices+list(range(start,end))

    try:
        span=doc.char_span(start,end,label=label,alignment_mode='strict')
    except:
        continue

    if span is None:
        err_data=str([start,end])+ " "+str(text)+'\n'
        file.write(err_data)

    else:
        ents.append(span)

    try:
        doc.ents=ents
        db.add(doc)
    except:
        pass
    return db

from sklearn.model_selection import train_test_split
train,test=train_test_split(cv_data,test_size=0.3)

len(train),len(test)

(140, 60)

file=open('error.txt','w')

db=get_spacy_doc(file,train)
db.to_disk('train_data.spacy')

db=get_spacy_doc(file,test)
db.to_disk('test_data.spacy')

100%|██████████| 140/140 [00:01<00:00, 86.60it/s]
100%|██████████| 60/60 [00:00<00:00, 91.56it/s]

len(db.tokens)

60

!python -m spacy train /content/Resume-Parser/data/training/config.cfg --output ./output --paths.train
[2023-09-30 07:33:14.035987: W tensorflow/compiler/tf2tensorrt/utils/py_utils.cc:38] TF-TRT Warning: Could not find TensorRT
✓ Created output directory: output
i Saving to output directory: output
i Using GPU: 0

===== Initializing pipeline =====
Downloading (...)lve/main/config.json: 100% 481/481 [00:00<00:00, 1.59MB/s]
Downloading (...)olve/main/vocab.json: 100% 899k/899k [00:00<00:00, 6.85MB/s]
Downloading (...)olve/main/merges.txt: 100% 456k/456k [00:00<00:00, 2.33MB/s]
Downloading (...)main/tokenizer.json: 100% 1.36M/1.36M [00:00<00:00, 19.6MB/s]
Downloading model.safetensors: 100% 499M/499M [00:01<00:00, 260MB/s]
Some weights of RobertaModel were not initialized from the model checkpoint at roberta-base and are newly initialized: ['roberta.pos']
You should probably TRAIN this model on a down-stream task to be able to use it for predictions and inference.
✓ Initialized pipeline

===== Training pipeline =====
i Pipeline: ['transformer', 'ner']
i Initial learn rate: 0.0
E   #      LOSS TRANS...  LOSS NER  ENTS_F  ENTS_P  ENTS_R  SCORE
E   ----

```

```
✓ Saved pipeline to output directory
output/model-last
```

```
#test1
doc=nlp('My name is Aditya . I am working in Microsoft. I have 2 years of experience')
for ent in doc.ents:
    print(ent.text," >>>>>>" , ent.label_)

    Aditya .      >>>>>> Name
    Microsoft    >>>>>> Companies worked at
```

```
!pip install PyMuPDF
```

```
doc=nlp(text)
```

```
for ent in doc.ents:
    print(ent.text," >>>>>>" , ent.label_)

Alice Clark >>>>>> Name
AI / Machine Learning Delhi, India >>>>>> Designation
Microsoft >>>>>> Companies worked at
Software Engineer >>>>>> Designation
Microsoft >>>>>> Companies worked at
Microsoft >>>>>> Companies worked at
Microsoft >>>>>> Companies worked at
Microsoft >>>>>> Companies worked at
Microsoft >>>>>> Companies worked at
Microsoft >>>>>> Companies worked at
Indian Institute of Technology >>>>>> College Name
Machine Learning, Natural Language Processing, and Big Data Handling ADDITIONAL INFORMATION Professional Skills • Excellent analyti
```