```
!pip install spacy_transformers
!pip install spacy
import spacy
from spacy.tokens import DocBin
from tqdm import tqdm
import ison
spacy.__version_
     '3.6.1'
!nvidia-smi
     Sat Sen 30 07:32:07 2023
       NVIDIA-SMI 525.105.17 Driver Version: 525.105.17 CUDA Version: 12.0
       GPU Name
                       Persistence-M Bus-Id
                                                      Disp.A | Volatile Uncorr. ECC |
       Fan Temp Perf Pwr:Usage/Cap| Memory-Usage | GPU-Util Compute M.
      _______
       Default
                                                                                   N/A
       Processes:
       GPU GI CI
                                                                             GPU Memory
                              PID Type Process name
              ID
                   ID
      No running processes found
!git clone https://github.com/mohanty-aditya/Resume-Parser.git
     Cloning into 'Resume-Parser'...
     remote: Enumerating objects: 13, done.
     remote: Counting objects: 100% (13/13), done.
     remote: Compressing objects: 100% (12/12), done.
     remote: Total 13 (delta 1), reused 0 (delta 0), pack-reused 0
     Receiving objects: 100% (13/13), 506.45 KiB | 1.86 MiB/s, done.
     Resolving deltas: 100% (1/1), done.
cv_data=json.load(open('/content/Resume-Parser/data/training/train_data.json','r'))
cv data[0]
     ['Govardhana K Senior Software Engineer Bengaluru, Karnataka, Karnataka - Email me on Indeed: indeed.com/r/Govardhana-K/
     b2de315d95905b68 Total IT experience 5 Years 6 Months Cloud Lending Solutions INC 4 Month • Salesforce Developer Oracle 5 Years 2
     Month • Core Java Developer Languages Core Java, Go Lang Oracle PL-SQL programming, Sales Force Developer with APEX. Designations
     & Promotions Willing to relocate: Anywhere WORK EXPERIENCE Senior Software Engineer Cloud Lending Solutions - Bangalore,
     Karnataka - January 2018 to Present Present Senior Consultant Oracle - Bangalore, Karnataka - November 2016 to December 2017
     Staff Consultant Oracle - Bangalore, Karnataka - January 2014 to October 2016 Associate Consultant Oracle - Bangalore,
     Karnataka - November 2012 to December 2013 EDUCATION B.E in Computer Science Engineering Adithya Institute of Technology
     Tamil Nadu September 2008 to June 2012 <a href="https://www.indeed.com/r/Govardhana-K/b2de315d95905b68?isid=rex-download&ikw=download-">https://www.indeed.com/r/Govardhana-K/b2de315d95905b68?isid=rex-download&ikw=download-</a>
     top&co=IN https://www.indeed.com/r/Govardhana-K/b2de315d95905b68?isid=rex-download&ikw=download-top&co=IN SKILLS APEX. (Less
     than 1 year), Data Structures (3 years), FLEXCUBE (5 years), Oracle (5 years), Algorithms (3 years) LINKS
     https://www.linkedin.com/in/govardhana-k-61024944/ ADDITIONAL INFORMATION Technical Proficiency: Languages: Core Java, Go Lang,
     Data Structures & Algorithms, Oracle PL-SQL programming, Sales Force with APEX. Tools: RADTool, Jdeveloper, NetBeans, Eclipse, SQL
     developer, PL/SQL Developer, WinSCP, Putty Web Technologies: JavaScript, XML, HTML, Webservice Operating Systems: Linux, Windows
     Version control system SVN & Git-Hub Databases: Oracle Middleware: Web logic, OC4J Product FLEXCUBE: Oracle FLEXCUBE Versions 10.x,
     11.x and 12.x <a href="https://www.linkedin.com/in/govardhana-k-61024944/">https://www.linkedin.com/in/govardhana-k-61024944/</a>,
      {'entities': [[1749, 1755, 'Companies worked at'],
        [1696, 1702, 'Companies worked at'],
[1417, 1423, 'Companies worked at'],
[1356, 1793, 'Skills'],
        [1306, 1795, Skills],
[1209, 1215, 'Companies worked at'],
[1136, 1247, 'Skills'],
[928, 932, 'Graduation Year'],
[858, 889, 'College Name'],
[821, 856, 'Degree'],
[727, 701, 'Graduation Year']
         [787, 791, 'Graduation Year'],
         [744, 750, 'Companies worked at'],
         [722, 742, 'Designation'],
         [658, 664, 'Companies worked at'],
         [640, 656, 'Designation'],
         [574, 580, 'Companies worked at'],
        [575, 572, 'Designation'],
[470, 493, 'Companies worked at'],
[444, 468, 'Designation'],
```

```
[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/utils/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 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
     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):
  nlp=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))
       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)
     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.poc
    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
                LOSS TRANS... LOSS NER ENTS_F ENTS_P ENTS_R SCORE
                                        -----
```

```
a
              a
                      2923.28 1593.73
                                          0 01
                                                   0 01
                                                          0.27
                                                                    a aa
                     114086.02 65357.28 26.60
34052.80 23294.04 48.82
      4
            200
                                                  58.11
                                                          17.25
                                                                    0.27
      8
            400
                                                  62.06
                                                          40.24
                                                                    0.49
                      6379.13 19530.21 51.58
                                                          43.58
      12
            600
                                                  63.18
                                                                    0.52
            800
                       2995.44 17403.41
                                           55.18
                                                   59.29
                                                           51.60
                                                                    0.55
                     12014.75 16619.83 57.65
      20
           1000
                                                  64.46
                                                           52.14
                                                                    0.58
                       2330.43 15214.68
      24
           1200
                                           57.20
                                                  60.63
                                                           54.14
                                                                   0.57
                      1003.63 14422.61 56.24
      28
           1400
                                                  62.77
                                                           50.94
                                                                   0.56
           1600
                       820.35 13794.32
9289.82 13666.11
                                                           52.41
      32
                                           56.89
                                                  62.22
                                                                   0.57
      36
           1800
                                          55.85
                                                  54.91
                                                           56.82
                                                                   0.56
           2000
                       673.12 13007.43
      40
                                          57.64
                                                  58 07
                                                           57.22
                                                                   0 58
                        311.82 12467.04
     44
           2200
                                          53.26
                                                  47.44
                                                          60.70
                                                                   0.53
     48
           2400
                       2563.41 11974.69
                                          53.91
                                                  48.16
                                                          61.23
                                                                   0.54
           2600
                        230.38 10900.64
                                           55.16
                                                  48.91
                                                          63.24
                                                                    0.55

√ Saved pipeline to output directory

     output/model-last
nlp=spacy.load('/content/output/model-best')
```

```
doc=nlp('My name is Aditya . I am working in Microsoft. I have 2 years of experience')
for ent in doc.ents:
                  >>>>>" , ent.label_)
 print(ent.text,"
     Aditva .
               >>>>> Name
     Microsoft
                >>>>> Companies worked at
#test 2 using real time resume
!pip install PyMuPDF
     Collecting PyMuPDF
       Downloading PyMuPDF-1.23.4-cp310-none-manylinux2014_x86_64.whl (4.3 MB)
                                                  - 4.3/4.3 MB 28.8 MB/s eta 0:00:00
     Collecting PyMuPDFb==1.23.3 (from PyMuPDF)
       Downloading PyMuPDFb-1.23.3-py3-none-manylinux2014_x86_64.manylinux_2_17_x86_64.whl (30.6 MB)
                                                  - 30.6/30.6 MB 34.5 MB/s eta 0:00:00
     Installing collected packages: PyMuPDFb, PyMuPDF
     Successfully installed PyMuPDF-1.23.4 PyMuPDFb-1.23.3
import sys, fitz
fname='/content/Alice Clark CV.pdf'
doc=fitz.open(fname)
text=' '
for page in doc:
 text=text+str(page.get_text())
```

' Alice Clark \nAI / Machine Learning \n \nDelhi, India Email me on Indeed \n• \n20+ years of experience in data handling, design, and development \n• \nData Warehouse: Data analysis, star/snow flake scema data modelling and design specific to \ndata warehousing and business intelligence $\n \cdot \n$ in Database: Experience in database designing, scalability, back-up and recovery, writing and \n zing SQL code and Stored Procedures, creating functions, views, triggers and indexes. \nCloud platform: Worked on Microsoft Azure c loud services like Document DB, SQL Azure, \nStream Analytics, Event hub, Power BI, Web Job, Web App, Power BI, Azure data lake \na $\verb|nalytics(U-SQL)| $$ \arrowvert - Bangalore, Karnataka \nJanuar - Bangalore, Banga$ y 2000 to Present \n1. Microsoft Rewards Live dashboards: \nDescription: - Microsoft rewards is loyalty program that rewards Users for browsing and shopping \nonline. Microsoft Rewards members can earn points when ...

```
text=' '.join(text.split())
```

text

text

'Alice Clark AI / Machine Learning Delhi, India Email me on Indeed • 20+ years of experience in data handling, design, and developm ent • Data Warehouse: Data analysis, star/snow flake scema data modelling and design specific to data warehousing and business inte lligence • Database: Experience in database designing, scalability, back-up and recovery, writing and optimizing SQL code and Store d Procedures, creating functions, views, triggers and indexes. Cloud platform: Worked on Microsoft Azure cloud services like Docume nt DB, SQL Azure, Stream Analytics, Event hub, Power BI, Web Job, Web App, Power BI, Azure data lake analytics(U-SQL) Willing to re locate anywhere WORK EXPERIENCE Software Engineer Microsoft - Bangalore, Karnataka January 2000 to Present 1. Microsoft Rewards Liv e dashboards: Description: - Microsoft rewards is loyalty program that rewards Users for browsing and shopping online. Microsoft Re wards members can earn points when searching with Bing, browsing with Microsoft Edge...

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
Indian Institute of Technology >>>>>> College Name
Machine Learning, Natural Language Processing, and Big Data Handling ADDITIONAL INFORMATION Professional Skills • Excellent analyti
```