PSAI

#PeopleSpace #NLP #doccano #spaCy

ZeroPage · 최민준

#PSAI

다빈치SW교육원
Da Vinci Software Education Institute

SW교육 | 산학·창업·국제협력 | SW가치확산 | 신청·예약 | 커뮤니티 | 기관 소개

해외인턴십 2022 PeopleSpace AI Summer 해외 인턴십 모집(추가 모집 및 기간 연장)

기업명	PeopleSpace		
진행기간	2022-07-18 ~ 2022-08-12	인원	15명
신청기간	2022-05-21 ~ 2022-06-07	진행상태	진행중
근무장소	미국		
첨부	2022 PSAI Summer Applicant Resume.docx (양식)2022 PSAI 참가확인서(부모님확약서).hwp 2022 해외 인턴십(People Space) 프로그램 개요.pdf		

#Business Model

- Advanced employee를 위한 resume 제작 assistant
- NLP (Natural Language Processing)

#NLP : Natural Language Processing

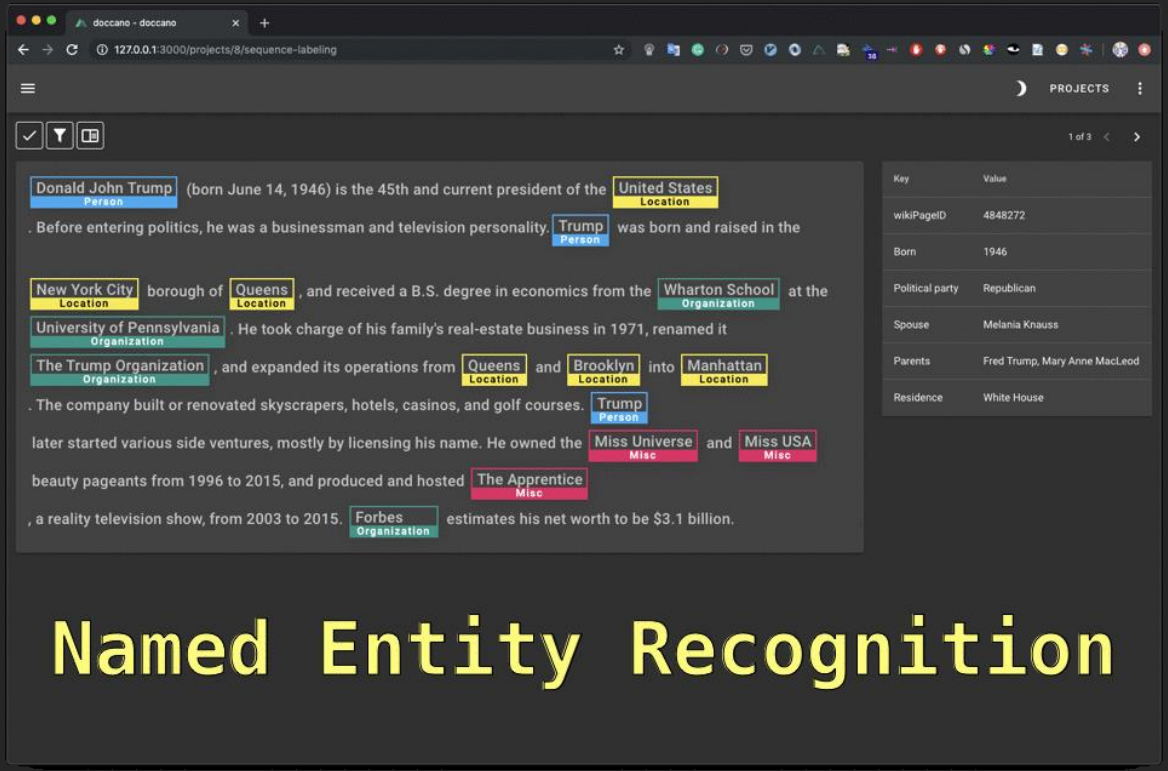
- 데이터 생성
- 인공지능 모델 학습

#Data

- Resume, Job Description from Kaggle
- Annotation with Doccano

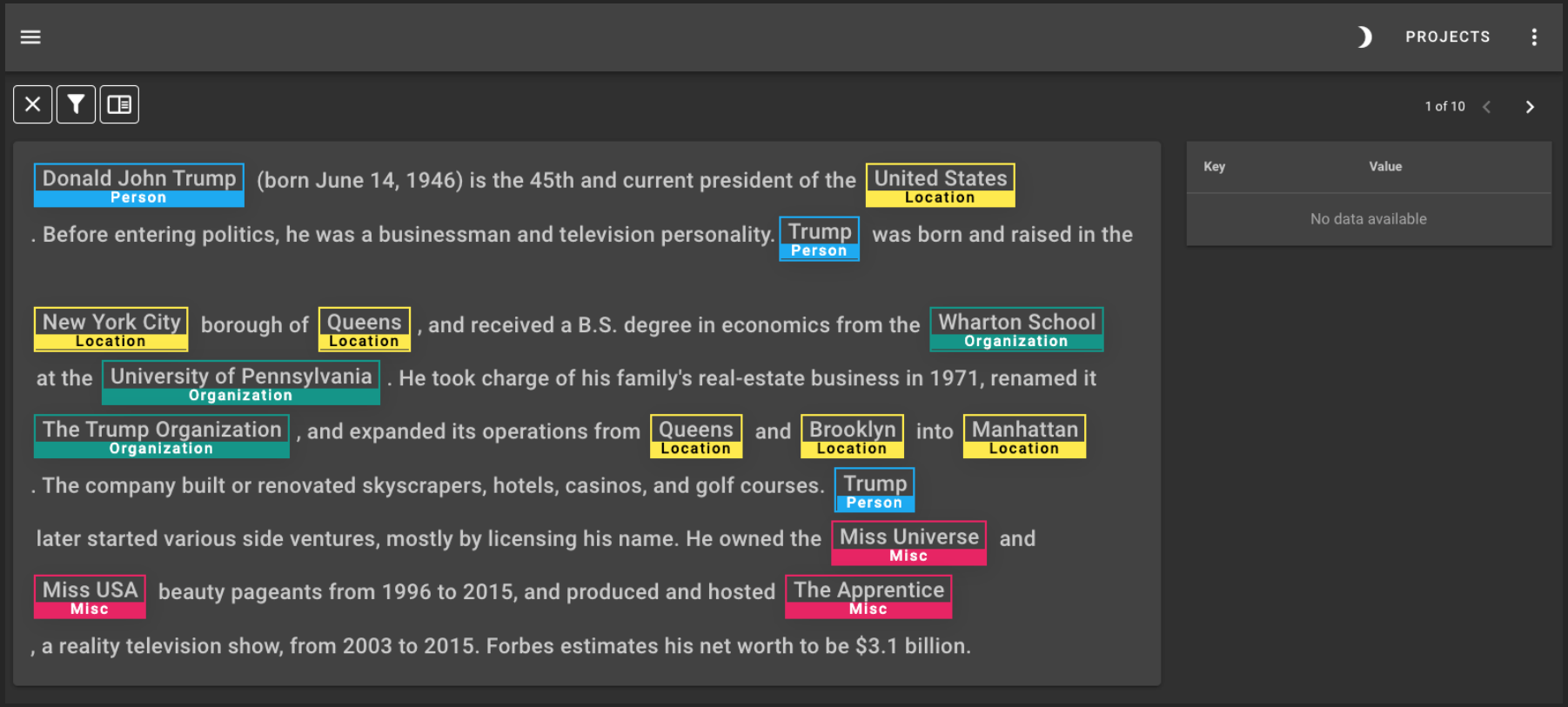
#Doccano

: Open source annotation tool for machine learning



#Doccano #NER

Resume 110개
Job Description 160개



#Doccano #NER


```
[
  {
    "id": 297,
    "text": "Web Developer Resume Template (Full Text Version)\nOlivier…,,
    "label": [
        [0, 13, "ability"],
        [623, 655, "work experience"],
        [671, 724, "work experience"],
        [1202, 1249, "work experience"],
        [1274, 1349, "work experience"],
        [1399, 1437, "work experience"],
        [1600, 1655, "award"],
        [1686, 1712, "major"],
        …
    ]
  }
]
```

#spaCy

“ 파이썬의 자연어처리를 위한 오픈 소스 기반 라이브러리 ”

#spaCy #Loss Function

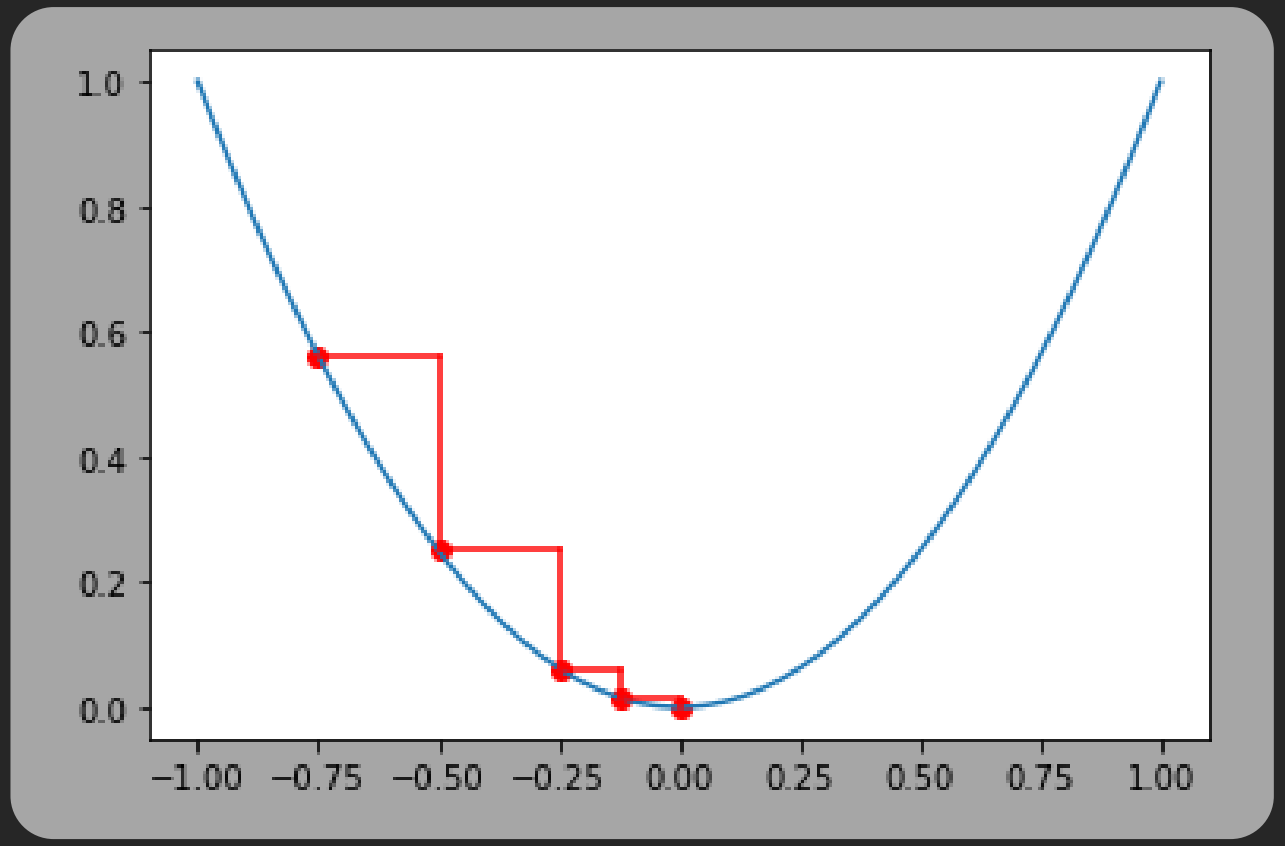
Java

skill | college | work experience | project

#spaCy #Optimizer

- Loss가 최소가 되도록 weight, bias 조정
- Gradient Descent, Momentum, SGD, etc

#spaCy #Gradient Descent



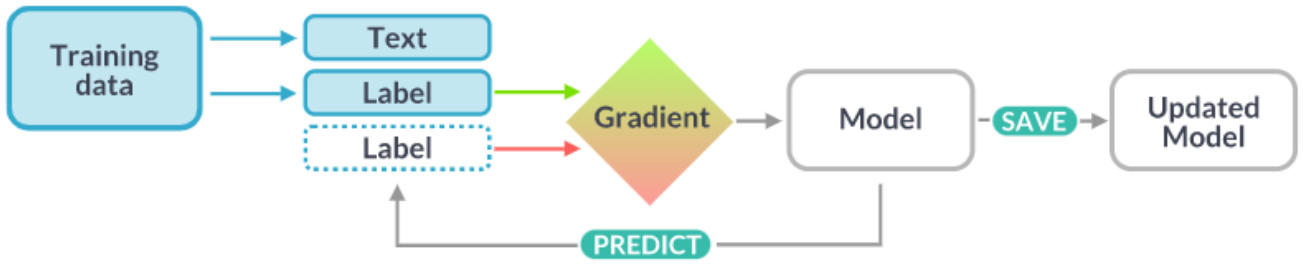
$$w = w - \alpha \frac{\delta L}{\delta w}$$

$$\mathrm{b} = \mathrm{b} - \alpha \frac{\delta L}{\delta b}$$

#spaCy

spaCy's tagger, parser, text categorizer and many other components are powered by **statistical models**. Every "decision" these components make – for example, which part-of-speech tag to assign, or whether a word is a named entity – is a **prediction** based on the model's current **weight values**. The weight values are estimated based on examples the model has seen during **training**. To train a model, you first need training data – examples of text, and the labels you want the model to predict. This could be a part-of-speech tag, a named entity or any other information.

Training is an iterative process in which the model's predictions are compared against the reference annotations in order to estimate the **gradient of the loss**. The gradient of the loss is then used to calculate the gradient of the weights through backpropagation. The gradients indicate how the weight values should be changed so that the model's predictions become more similar to the reference labels over time.



#spaCy

- spaCy는 이미 trained model이 있음
- annotated data를 넣어주면 자체 model로 분석을 함.
- Input data와 자체 model이 분석한 차이를 비교. (loss 파악)
- gradient와 backpropagation을 통해 model에서 사용하는 w, b 조정

#spaCy #Result

Job Description

```
job description num 1 test
Django Developer  ->>>>>>>  job
Self managing  ->>>>>>>  required skills
1+ years  ->>>>>>>  required year
Python software development  ->>>>>>>  required skills
2+ years  ->>>>>>>  required year
System Integration using XML/REST  ->>>>>>>  required skills
Web Frameworks  ->>>>>>>  required skills
GIT  ->>>>>>>  preferred skills
Frontend skills  ->>>>>>>  required skills
HTML  ->>>>>>>  required skills
CSS  ->>>>>>>  required skills
Javascript  ->>>>>>>  required skills
JQuery  ->>>>>>>  required skills
B.Tech/B.E.  ->>>>>>>  preferred skills
Computers  ->>>>>>>  preferred skills
Electrical, Electronics/Telecommunication  ->>>>>>>  preferred skills
B.Sc  ->>>>>>>  preferred skills
- Computers  ->>>>>>>  preferred skills
```

Resume

```
resume num 1 test
Bachelor Of Technology in  ->>>>>>>  degree
Engineering And Technology  ->>>>>>>  major
Rajasthan Technical University  ->>>>>>>  college
ASP.NET  ->>>>>>>  skill
C  ->>>>>>>  skill
• Operating system:-  ->>>>>>>  work experience
• Language Known:-  ->>>>>>>  work experience
C  ->>>>>>>  skill
• Concepts Known:-  ->>>>>>>  work experience
DBMS  ->>>>>>>  skill
Computer architecture  ->>>>>>>  skill
DSA  ->>>>>>>  skill
• Front End  ->>>>>>>  work experience
HTML  ->>>>>>>  skill
CSS  ->>>>>>>  skill
XML  ->>>>>>>  skill
AJAX  ->>>>>>>  skill
• Web Development:-  ->>>>>>>  work experience
ASP.NET  ->>>>>>>  skill
• RDBMS:- SQL  ->>>>>>>  work experience
```

#spaCy #Result

Job Description

```
job description num 1 test
Django Developer  ->>>>>>>  job
Self managing  ->>>>>>>  required skills
1+ years  ->>>>>>>  required year
Python software development  ->>>>>>>  required skills
2+ years  ->>>>>>>  required year
System Integration using XML/REST  ->>>>>>>  required skill
Web Frameworks  ->>>>>>>  required skills
GIT  ->>>>>>>  preferred skills
Frontend skills  ->>>>>>>  required skills
HTML  ->>>>>>>  required skills
CSS  ->>>>>>>  required skills
Javascript  ->>>>>>>  required skills
JQuery  ->>>>>>>  required skills
B.Tech/B.E.  ->>>>>>>  preferred skills
Computers  ->>>>>>>  preferred skills
Electrical, Electronics/Telecommunication  ->>>>>>>  preferred skills
B.Sc  ->>>>>>>  preferred skills
- Computers  ->>>>>>>  preferred skills
```

Required Technical Experience:

1+ years of Python software development , 2+ years in any modern language.

System Integration using XML/REST APIs

Software architecture patterns

Web Frameworks and template languages

GIT version control

MySQL Database

Message based / Asynchronous processing systems

Deployment to Linux

Flask / Django

#spaCy #Limits

- Annotated Data 너무 주관적
- spaCy model 학습과정 알 수 없음
- 단어의 위치로 파악 X, Annotation 빈도

감사합니다 :)

•

Any Questions?