

FashionNER

Fashion Named Entity Recognition
by Ceyda Cinarel

05 June 2021

About Me

Ceyda Çınarel



 Contributing to Open Source



Mathematics
Computer Engineering



서울대학교

Computer Science & Engineering



2015 ↓

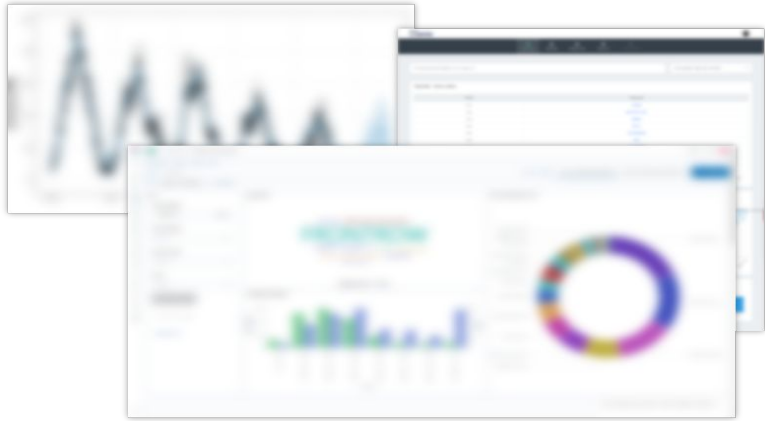


AI Researcher (NLP)

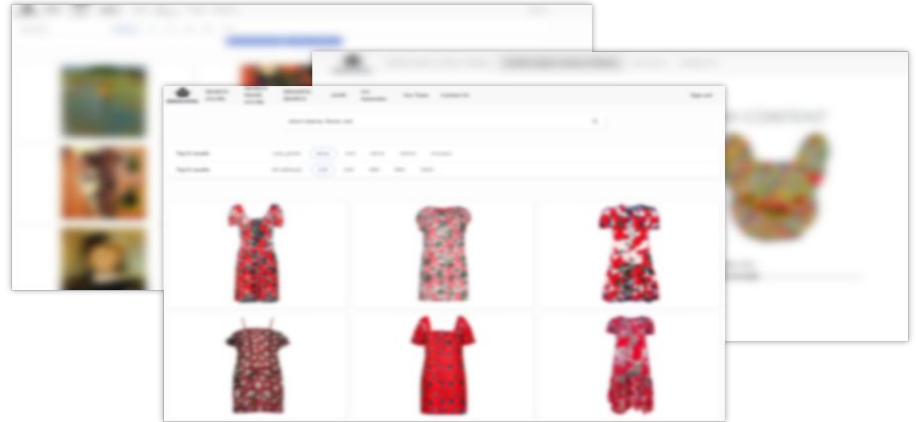


Senior Researcher & Engineer

Attention
Embeddings
Color
Augmentation
GAN
Style Transfer
Transformers
NLP ❤️ CV
Adversarial
Multimodal
QA
NER

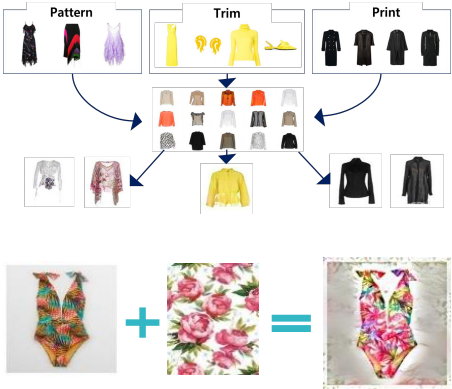
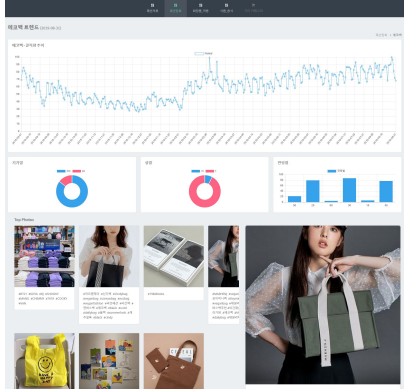


[Fashion Data Analysis Solutions]

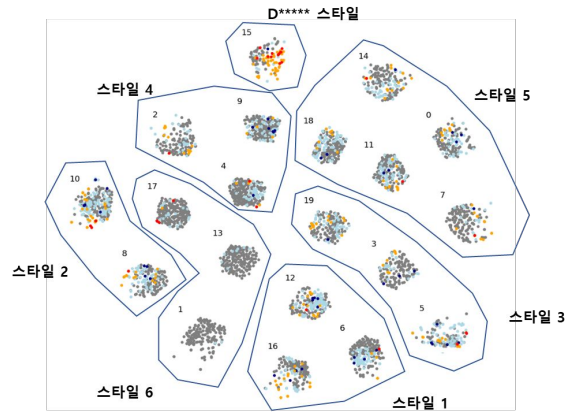
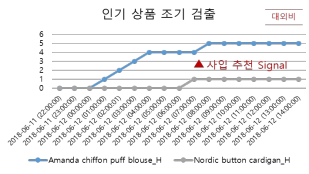


[Design AI ★]

- Multimodal Search Solutions
- Data Analysis & Prediction
 - Trends / Reviews ...
- Style Transfer
- GANs



Category	ProductID	BrandID	CategoryID	ColorID	Material
Brand	PHILIPHEIN	SHANGHAI	CHINA	RED	NYLON
Product	PHILIPHEIN	PHILIPHEIN	PHILIPHEIN	PHILIPHEIN	PHILIPHEIN
Pattern	PHILIPHEIN	PHILIPHEIN	PHILIPHEIN	PHILIPHEIN	PHILIPHEIN
Color	PHILIPHEIN	PHILIPHEIN	PHILIPHEIN	PHILIPHEIN	PHILIPHEIN
Material	PHILIPHEIN	PHILIPHEIN	PHILIPHEIN	PHILIPHEIN	PHILIPHEIN
Material	PHILIPHEIN	PHILIPHEIN	PHILIPHEIN	PHILIPHEIN	PHILIPHEIN



스타일별 대표 이미지



DESIGNNOVEL x Partnerships



ppb studios



SHINSEGAE INTERNATIONAL

COMODO
STUDIO



E·LAND RETAIL LTD.
(주)이랜드리테일

 landmall

LOTTE ON
DEMONZ



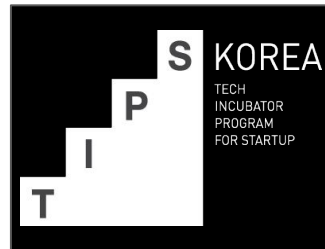
DESIGNNOVEL x Sponsors

Google
for
Startups



Microsoft
for Startups

 **NVIDIA**
INCEPTION PROGRAM



 *National IT Industry Promotion Agency*

Overview

- Problem Definition
- Data
 - Collection
 - Annotation
 - Processing
- Modelling
 - Train
 - Evaluate
- Deploy
 - Server
 - UI



AI Memes for Artificially Intelligent Teens
@ai_memes



Machine learning pipelines



1:45 AM · Apr 15, 2021 · Twitter Web App

9,119 Retweets 1,985 Quote Tweets 34.2K Likes

Overview

- Problem Definition
- Data
 - Collection
 - Annotation
 - Processing
- Modelling
 - Train
 - Evaluate
- Deploy
 - Server
 - UI



AI Memes for Artificially Intelligent Teens
@ai_memes



Machine learning pipelines



1:45 AM · Apr 15, 2021 · Twitter Web App

9,119 Retweets 1,985 Quote Tweets 34.2K Likes

What's an Entity anyway?

Named Entity:

Person, Location, Organization, Countries

What's an Entity anyway?

Named Entity:

Person, Location, Organization, Countries

Expanded to:

Products, Events

Numeric: Date, Time, Quantity, Ordinal, Money

What's an Entity anyway?

Named Entity:

Person, Location, Organization, Countries

Expanded to:

Products, Events

Numeric: Date, Time, Quantity, Ordinal, Money

Famous datasets:

CoNLL, WNUT-NER ...

What's an Entity anyway?

Named Entity:

Person, Location, Organization, Countries

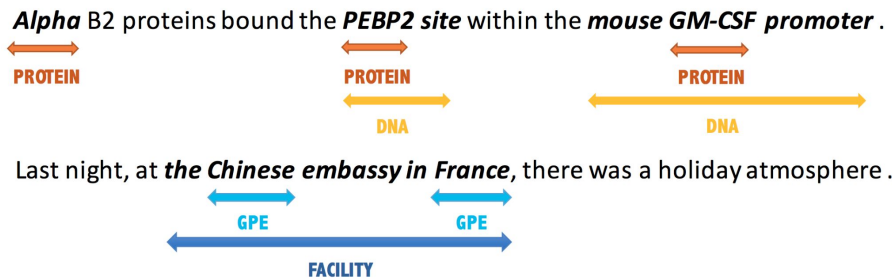
Expanded to:

Products, Events

Numeric: Date, Time, Quantity, Ordinal, Money

Famous datasets:

CoNLL, WNUT-NER ...



Generalized as

Token Classification / Sequence Labeling

What's an Entity anyway?

Tokens	IO	BIO	BMEWO	BMEWO+
Yesterday	O	O	O	BOS_O
afternoon	O	O	O	O
,	O	O	O	O_PER
John	I_PER	B_PER	B_PER	B_PER
J	I_PER	I_PER	M_PER	M_PER
.	I_PER	I_PER	M_PER	M_PER
Smith	I_PER	I_PER	E_PER	E_PER
traveled	O	O	O	PER_O
to	O	O	O	O_LOC
Washington	I_LOC	B_LOC	W_LOC	W_LOC
.	O	O	O	O_EOS

Generalized as

Token Classification / Sequence Labeling

What's an Entity anyway?

Tokens	IO	BIO	BMEWO	BMEWO+
Yesterday	O	O	O	BOS_O
afternoon	O	O	O	O
,	O	O	O	O_PER
John	I_PER	B_PER	B_PER	B_PER
J	I_PER	I_PER	M_PER	M_PER
.	I_PER	I_PER	M_PER	M_PER
Smith	I_PER	I_PER	E_PER	E_PER
traveled	O	O	O	PER_O
to	O	O	O	O_LOC
Washington	I_LOC	B_LOC	W_LOC	W_LOC
.	O	O	O	O_EOS

BIO - tagging format
The most commonly used format

There are between ups/downs
between different formats

Generalized as

Token Classification / Sequence Labeling



Balmain

[Brand Name](#)

short asymmetric ruffled leather skirt

[Title / Short
Description](#)

Your outfit could do with a rebellious design from Balmain. Crafted from lambskin, this miniskirt features an asymmetric hem with ruffled detailing that will make you stand out for all the right reasons. Dare to be yourself.

[Detailed Description](#)



Balmain

Brand Name

short asymmetric ruffled leather skirt

Title / Short
Description

Your outfit could do with a rebellious design from Balmain. Crafted from lambskin, this miniskirt features an asymmetric hem with ruffled detailing that will make you stand out for all the right reasons. Dare to be yourself.

Detailed Description



Balmain

Brand Name

short asymmetric ruffled leather skirt

Title / Short
Description

Your outfit could do with a rebellious design from Balmain. Crafted from lambskin, this miniskirt features an asymmetric hem with ruffled detailing that will make you stand out for all the right reasons. Dare to be yourself.

Detailed Description

Highlights

- black
- lambskin
- high waist
- signature B logo plaque
- rear zip fastening
- ruffled detailing
- asymmetric hem



Keywords

Related Research

flair

Back in 2000, **People Magazine** **PUBLISHER** highlighted **Prince Williams'** **PERSON** style who at the time was a little more fashion-conscious, even making fashion statements at times.

Now-a-days the prince mainly wears **navy** **COLOR** suits **ITEM** (sometimes **double-breasted** **DESIGN**), **light blue** **COLOR** button-ups **ITEM** with **classic** **LOOK** **pointed** **DESIGN** collars **PART**, and **burgundy** **COLOR** ties **ITEM**.

But who knows what the future holds ...

Duchess Kate **PERSON** did wear an **Alexander McQueen** **BRAND** dress **ITEM** to the **wedding** **OCCASION** in the **fall of 2017** **SEASON**.

Related Research

flair

Back in 2000, **People Magazine** PUBLISHER highlighted **Prince Williams'** PERSON style who at the time was a little more fashion-conscious, even making fashion statements at times.

Now-a-days the prince mainly wears **navy** COLOR suits ITEM (sometimes **double-breasted** DESIGN), **light blue** COLOR button-ups ITEM with **classic** LOOK **pointed** DESIGN collars PART, and **burgundy** COLOR ties ITEM.

But who knows what the future holds ...

Duchess Kate PERSON did wear an **Alexander McQueen** BRAND dress ITEM to the **wedding** OCCASION in the **fall of 2017** SEASON.

alanakbik commented on Dec 28, 2018

Collaborator

Hi @KeepingItClassy - this screenshot is from a "fashion" NER model that is currently proprietary to Zalando, so unfortunately we don't currently bundle it with the Flair release.

(But we do include general NER models such as the Ontonotes model which has a lot of classes you might find useful, as well as our frame detection model.)

Should you be interested in adding a frame-semantic parser to Flair we'd be very happy! Something like a FrameNet or PropBank parser has been on our list for a while, but we haven't yet gotten around to working on this.

👍 2

 zalando

2018

Related Research

flair

Back in 2000, **People Magazine** **PUBLISHER** highlighted **Prince Williams'** **PERSON** style who at the time was a little more fashion-conscious, even making fashion statements at times.

Now-a-days the prince mainly wears **navy** **COLOR** suits **ITEM** (sometimes **double-breasted** **DESIGN**), **light blue** **COLOR** button-ups **ITEM** with **classic** **LOOK** **pointed** **DESIGN** collars **PART**, and **burgundy** **COLOR** ties **ITEM**.

But who knows what the future holds ...

Duchess Kate **PERSON** did wear an **Alexander McQueen** **BRAND** dress **ITEM** to the **wedding** **OCCASION** in the **fall of 2017** **SEASON**.



(*'For', 'O'*), (*'Summer', 'O'*), (*'2013', 'O'*), (*'Supreme', 'O'*), (*'presents', 'O'*), (*'a', 'O'*), (*'small', 'O'*), (*'Wackies', 'O'*), (*'capsule', 'O'*), (*'collection', 'O'*), (*','*), (*'O'*), (*'consisting', 'O'*), (*'of', 'O'*), (*'four', 'O'*), (*'graphic', 'ITEM'*), (*'T-shirts', 'ITEM'*), (*','*), (*'O'*), (*'a', 'O'*), (*'tank', 'ITEM'*), (*'top', 'ITEM'*), (*','*), (*'O'*), (*'and', 'O'*), (*'a', 'O'*), (*'snapback', 'ITEM'*), (*'cap', 'ITEM'*)

fashion **BRAIN** D2.1 – Named Entity Recognition and Linking Methods

7

alanakbik commented on Dec 28, 2018

Collaborator

Hi @KeepingItClassy - this screenshot is from a "fashion" NER model that is currently proprietary to Zalando, so unfortunately we don't currently bundle it with the Flair release.

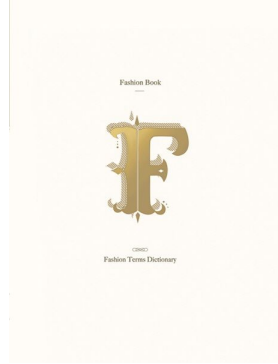
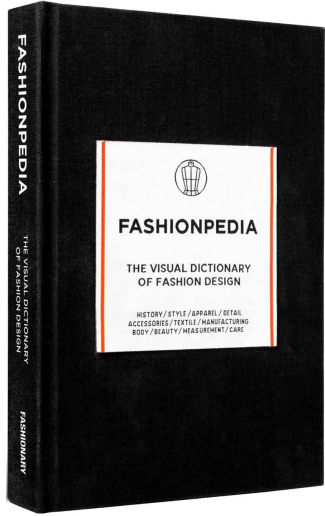
(But we do include general NER models such as the Ontonotes model which has a lot of classes you might find useful, as well as our frame detection model.)

Should you be interested in adding a frame-semantic parser to Flair we'd be very happy! Something like a FrameNet or PropBank parser has been on our list for a while, but we haven't yet gotten around to working on this.



2

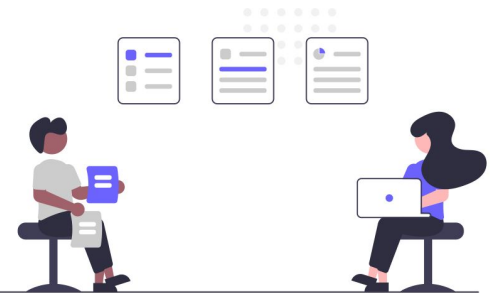
2018



+Online Resources

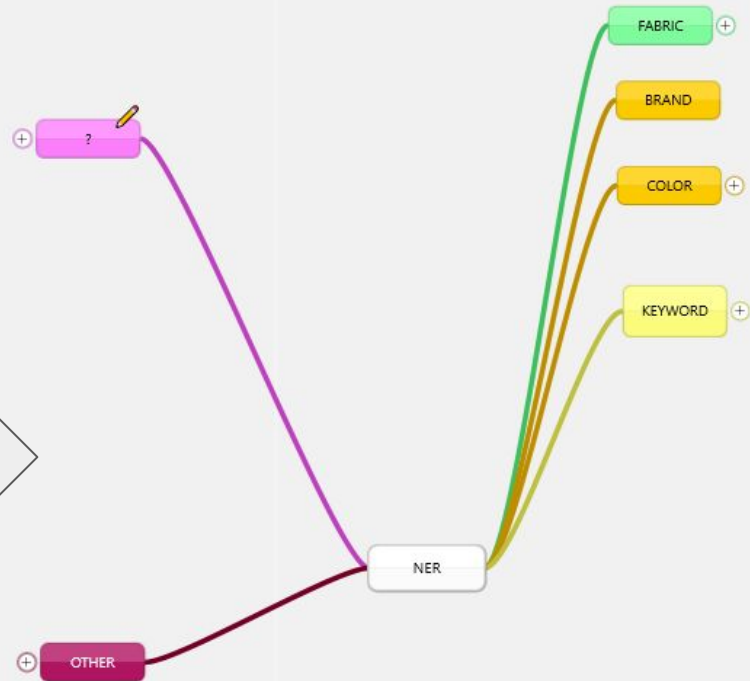
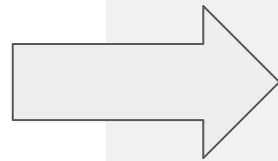
Fashion
Taxonomy/Dictionary
(Patterns)

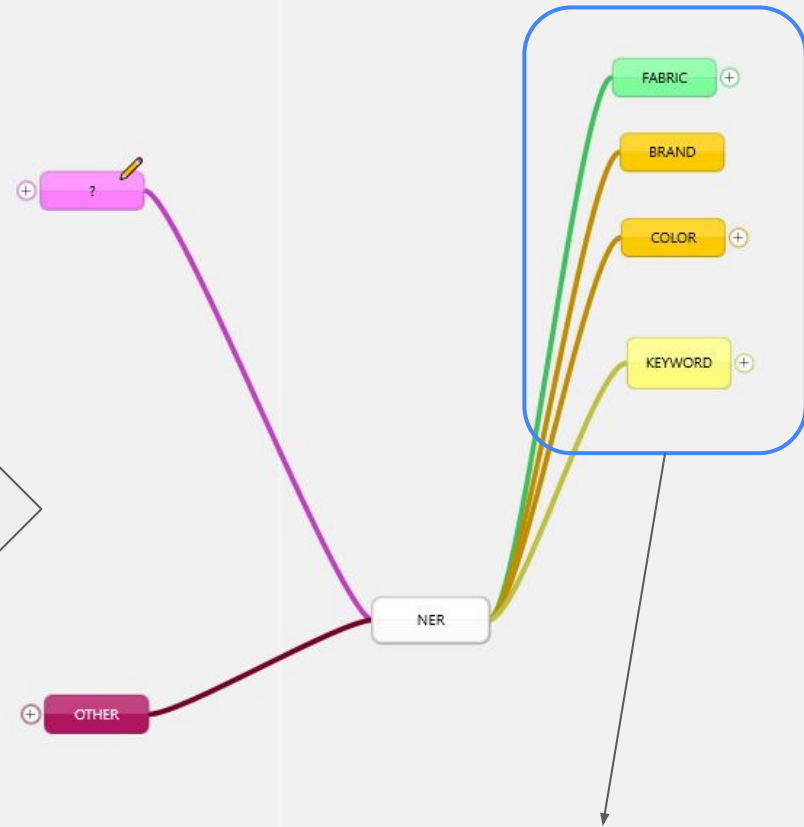
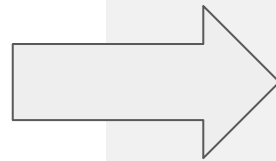
In House
Fashion Expert



A word cloud of textile-related terms. The words are arranged in a roughly triangular shape, with 'Brand' and 'Color' being the largest and most prominent. Other significant words include 'Textile', 'Category', 'Occasion', 'Pattern', and 'Composition'. Smaller words include 'Size', 'Attribute', 'Designer', 'Made in', 'Use', 'Age', 'Status', 'Length', 'Detail', 'Fit', 'Season', and 'Line'. The colors used are shades of blue, green, and purple.

Size
Occasion
Attribute
Designer
Brand
Pattern
Composition
Made in
Textile
Color
Use
Age
Status
Length
Category
Season
Detail
Fit
Line





Picked types of entities we are most interested in (for project part 1)

Why FashionNER

- Dual Meaning



fabric



item



color

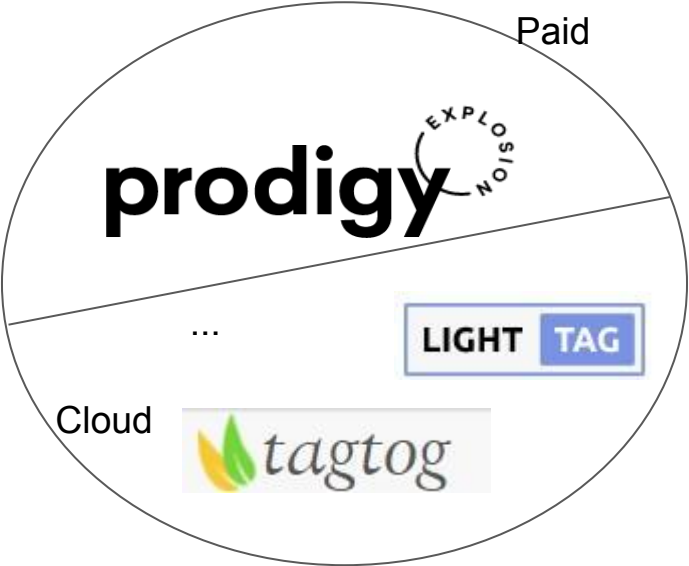
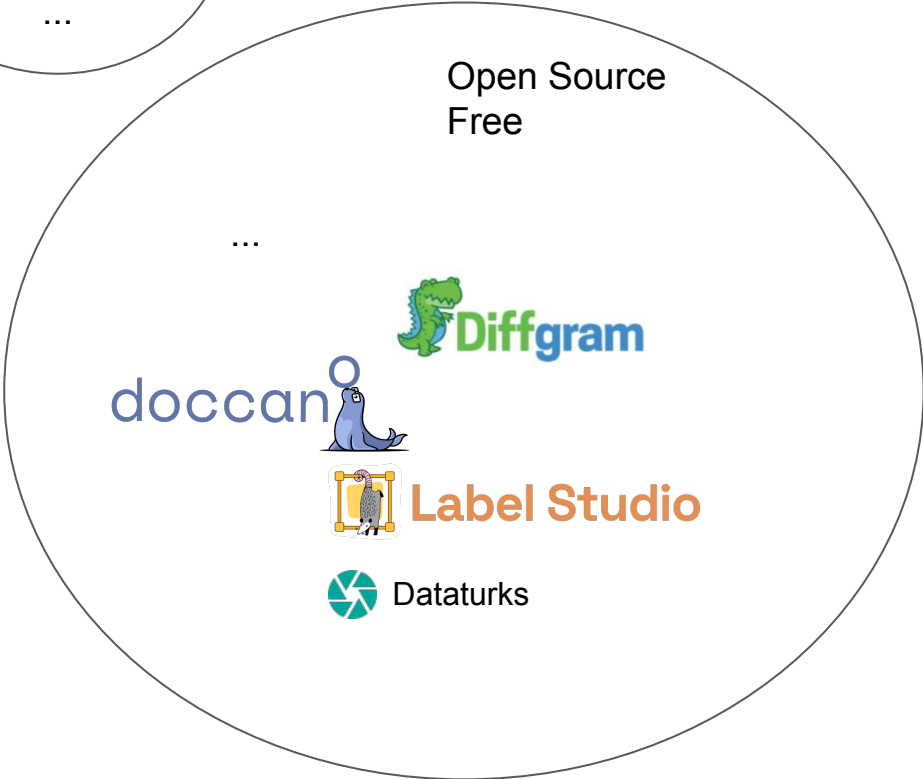
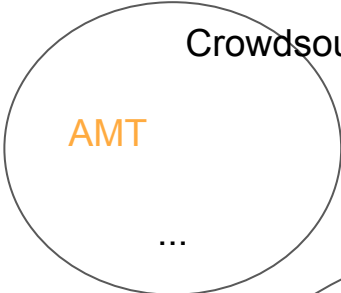
Off-WhiteTM

brand

- Expanding Vocab

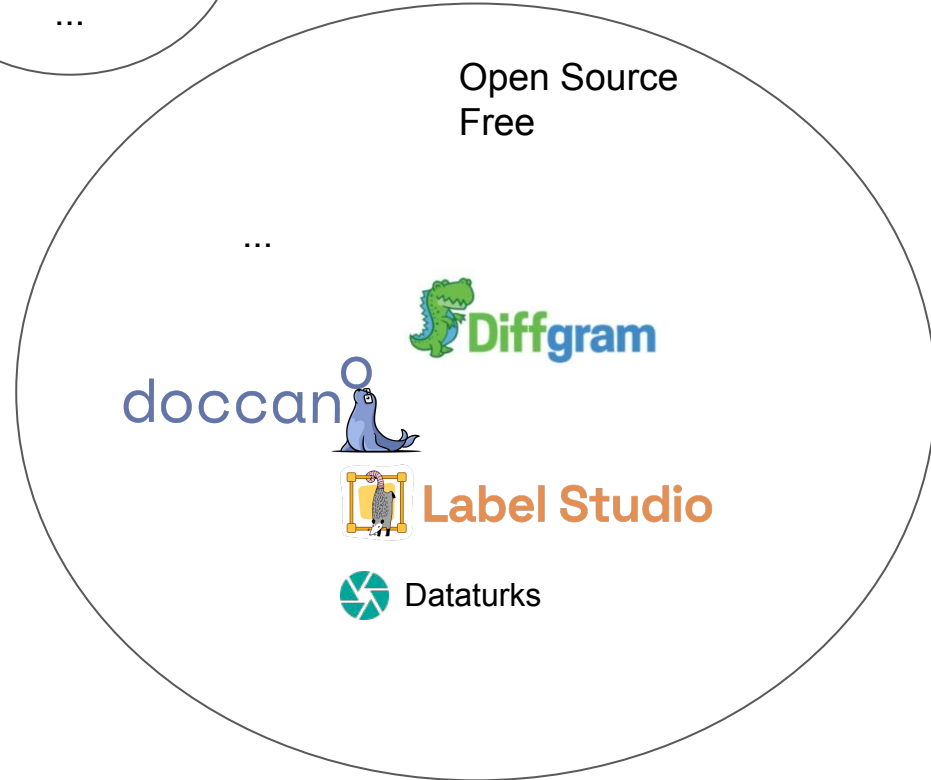
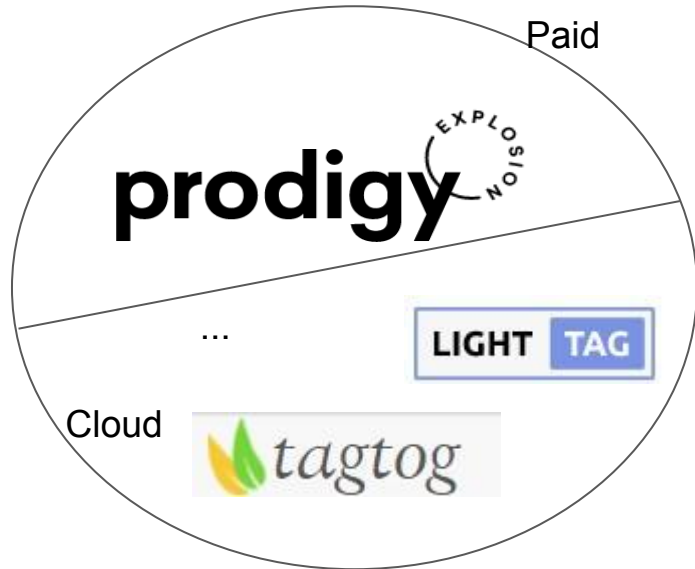
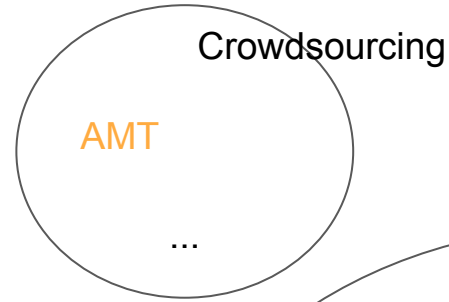
Context is important for entity resolution !

Data Annotation



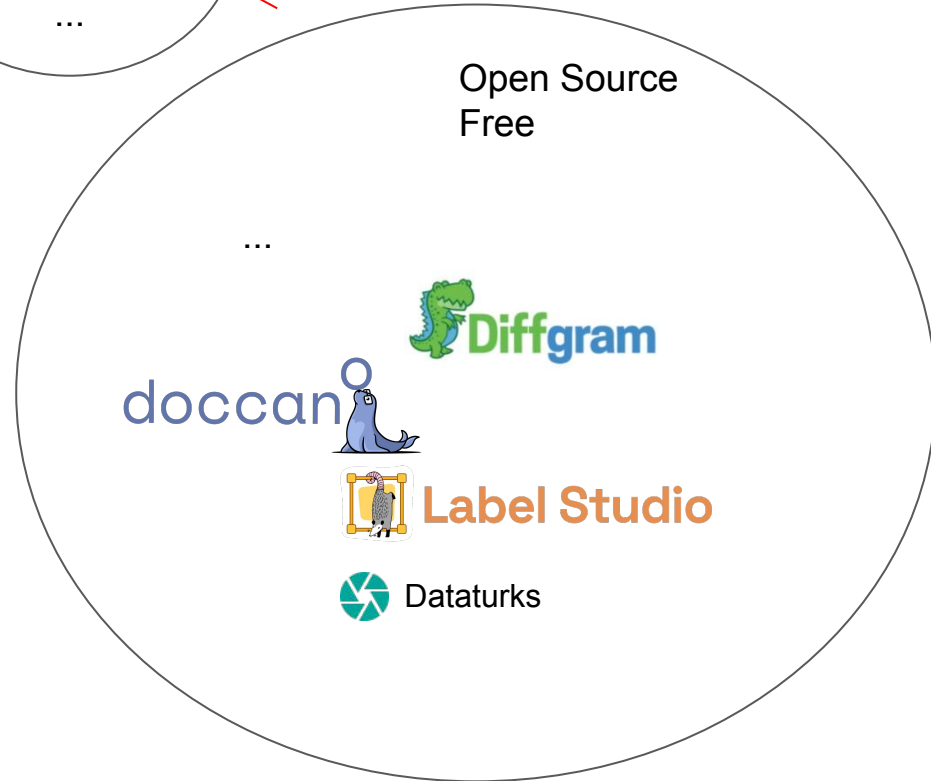
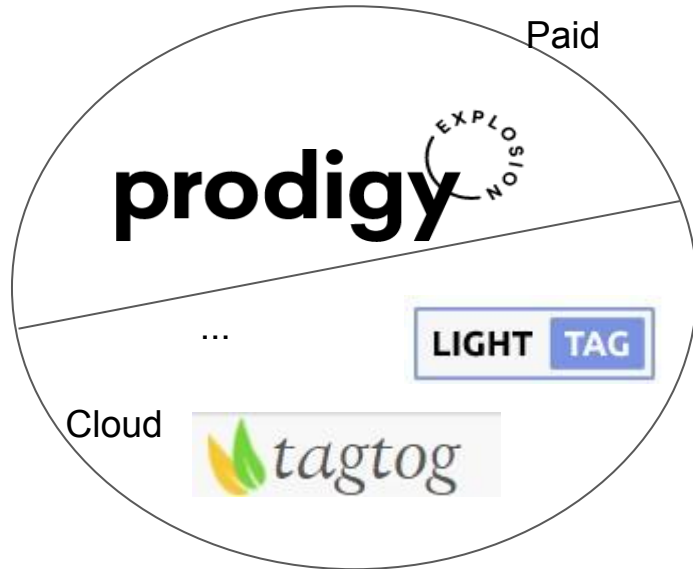
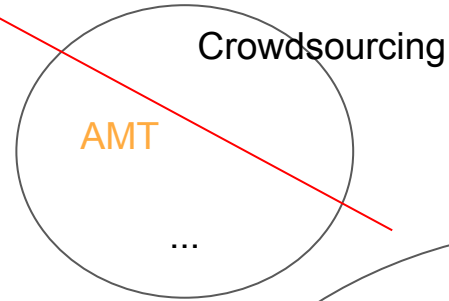
Data Annotation

- Who are your annotators?
 - Experts
 - Annotator guide/training



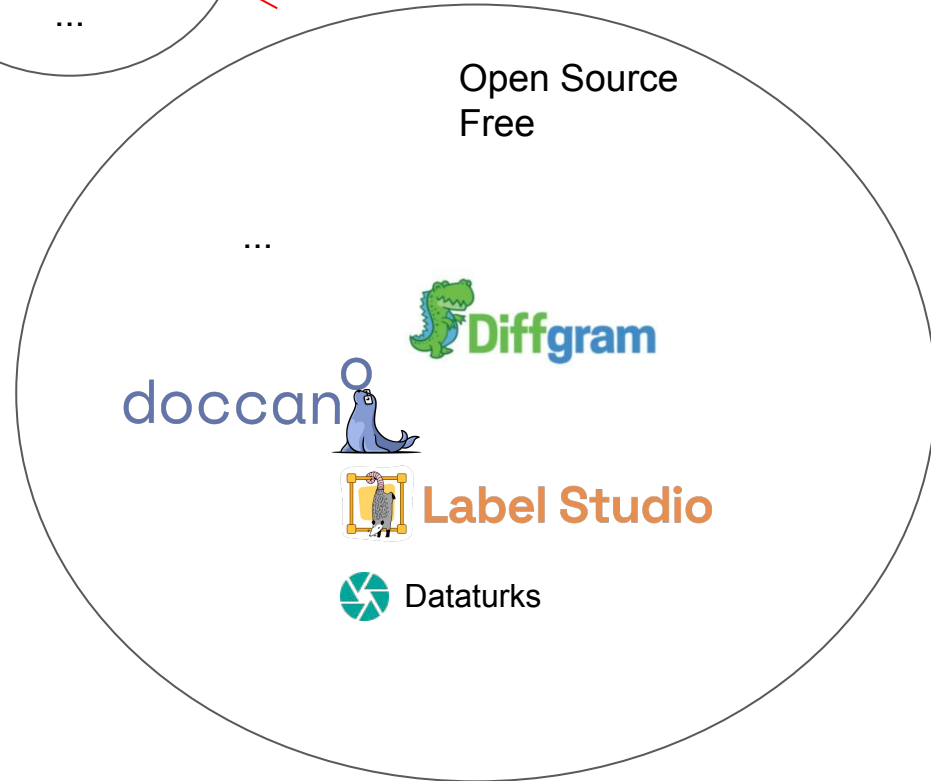
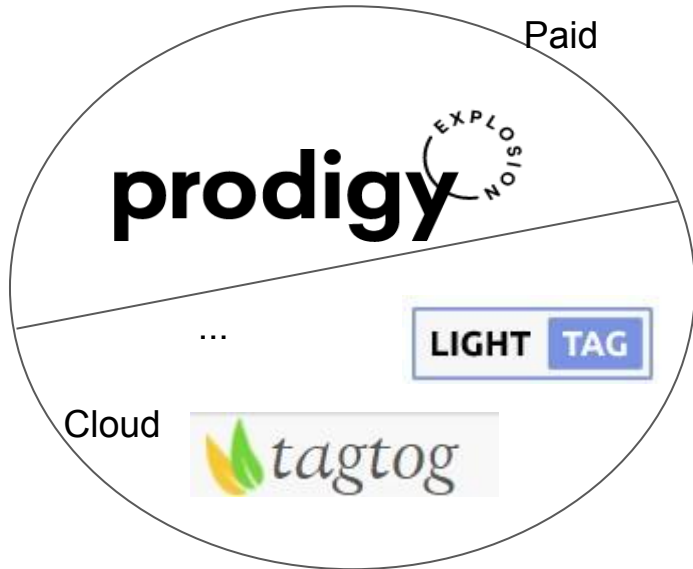
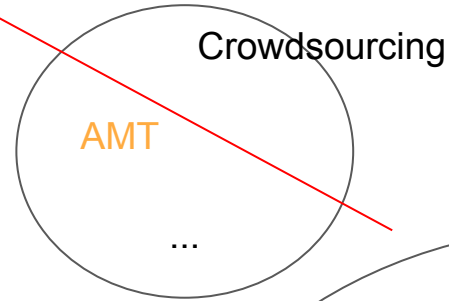
Data Annotation

- Who are your annotators?
- Data privacy?
 - Self-hosted options
 - Pipeline access



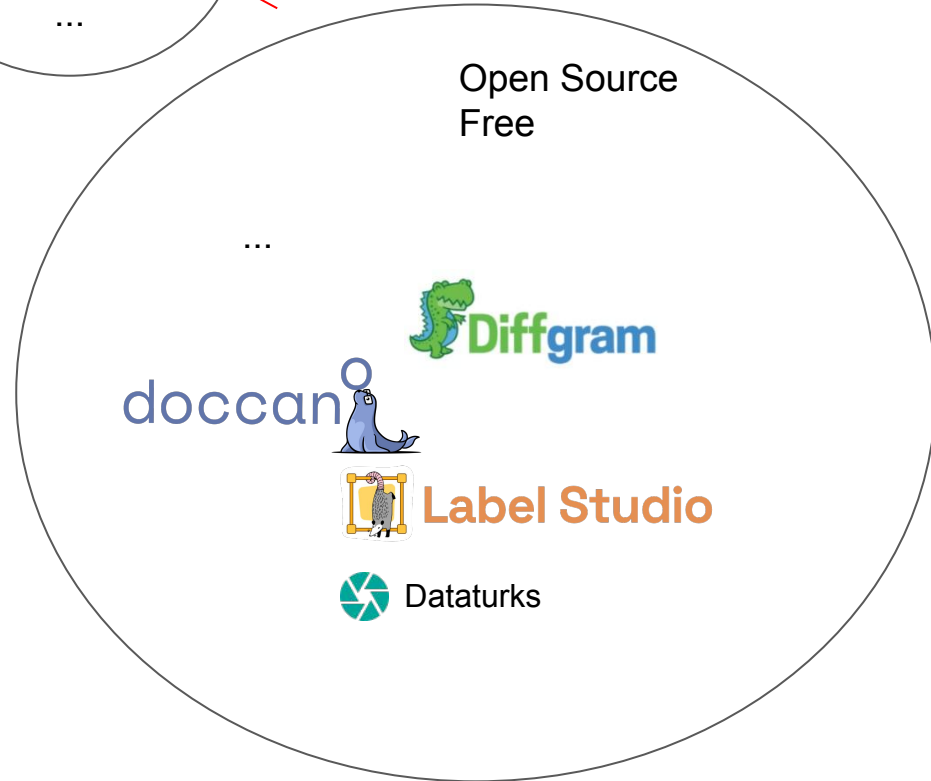
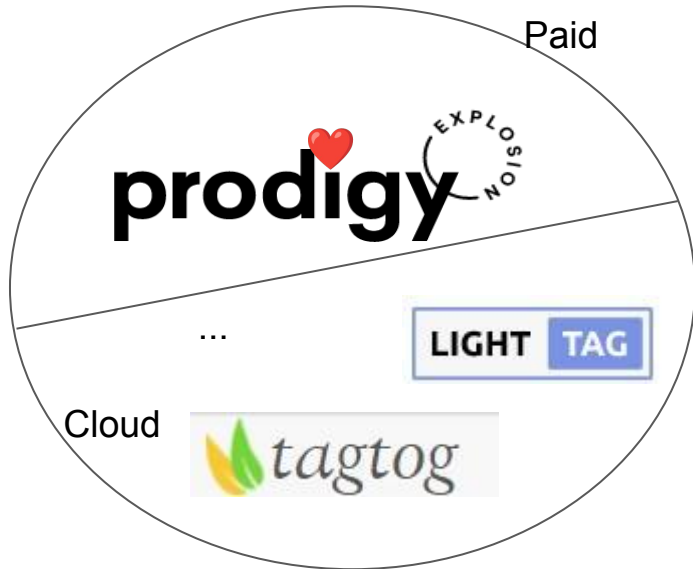
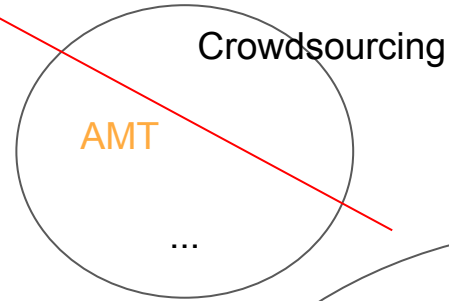
Data Annotation

- Who are your annotators?
- Data privacy?
- UI comfort / Ease of use?



Data Annotation

- Who are your annotators?
- Data privacy?
- UI comfort / Ease of use?



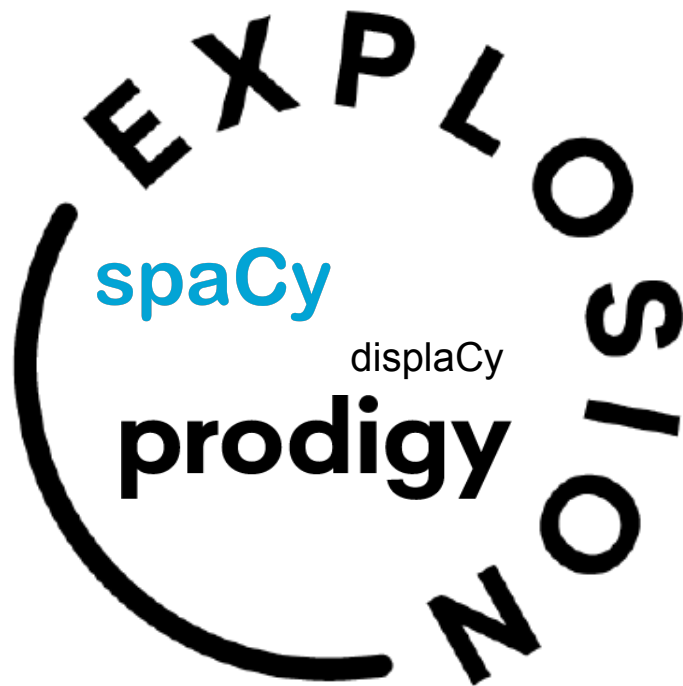
Why

Prodigy

- Multiple Annotation Types (CV+NLP+Audio)
- Python API pluggable custom annotation scripts
 - Streams / Callbacks
- Pretty UI
- Support

Format compatibility across tools makes life easier on the researcher 😊~

- + Warm fuzzy feeling you get for supporting spaCy development



Data Ingress



Annotation



Some Processing?



Dataset

- Provide Guidelines for Annotators

Data Ingress



Annotation

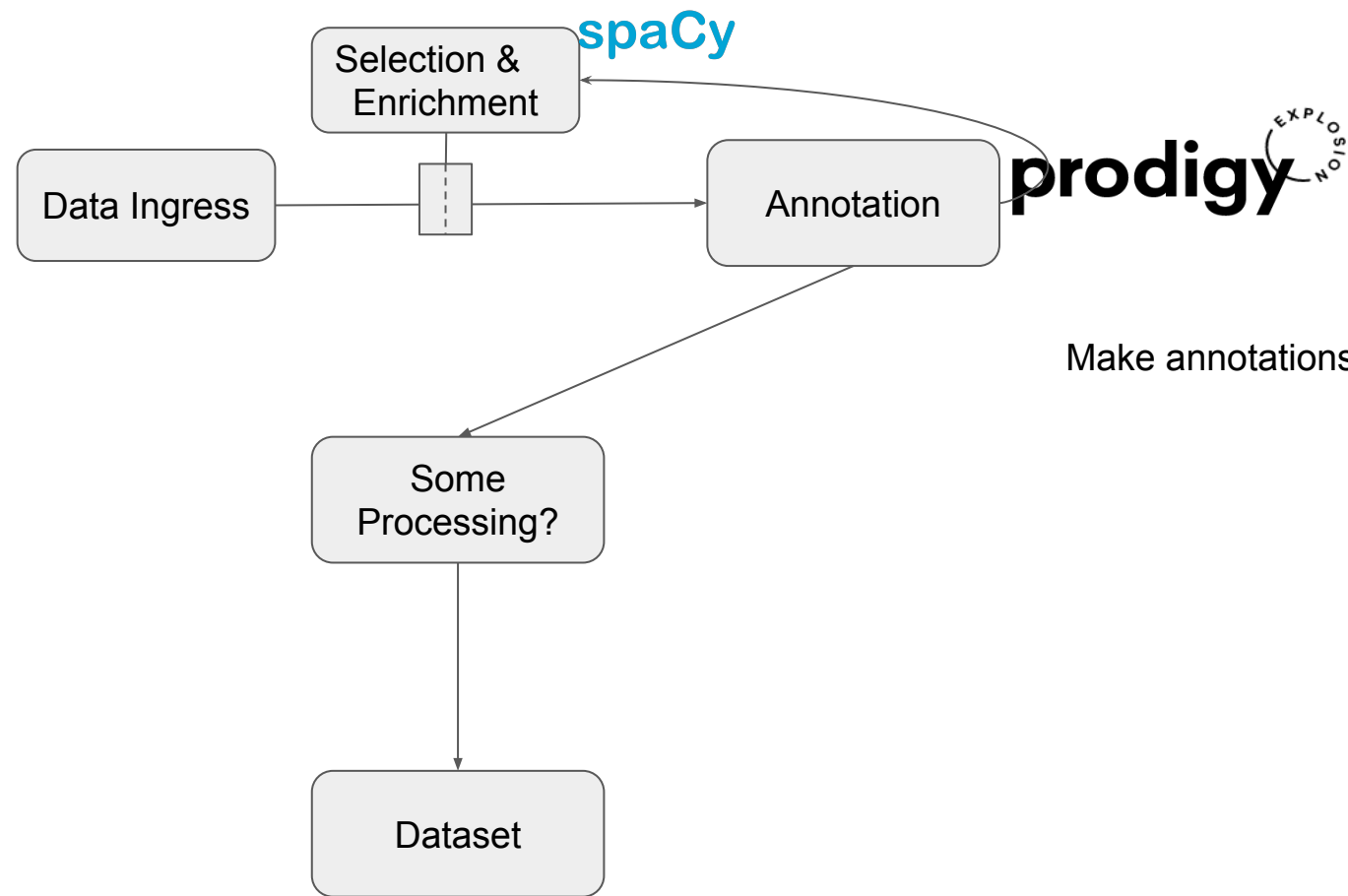


Some Processing?

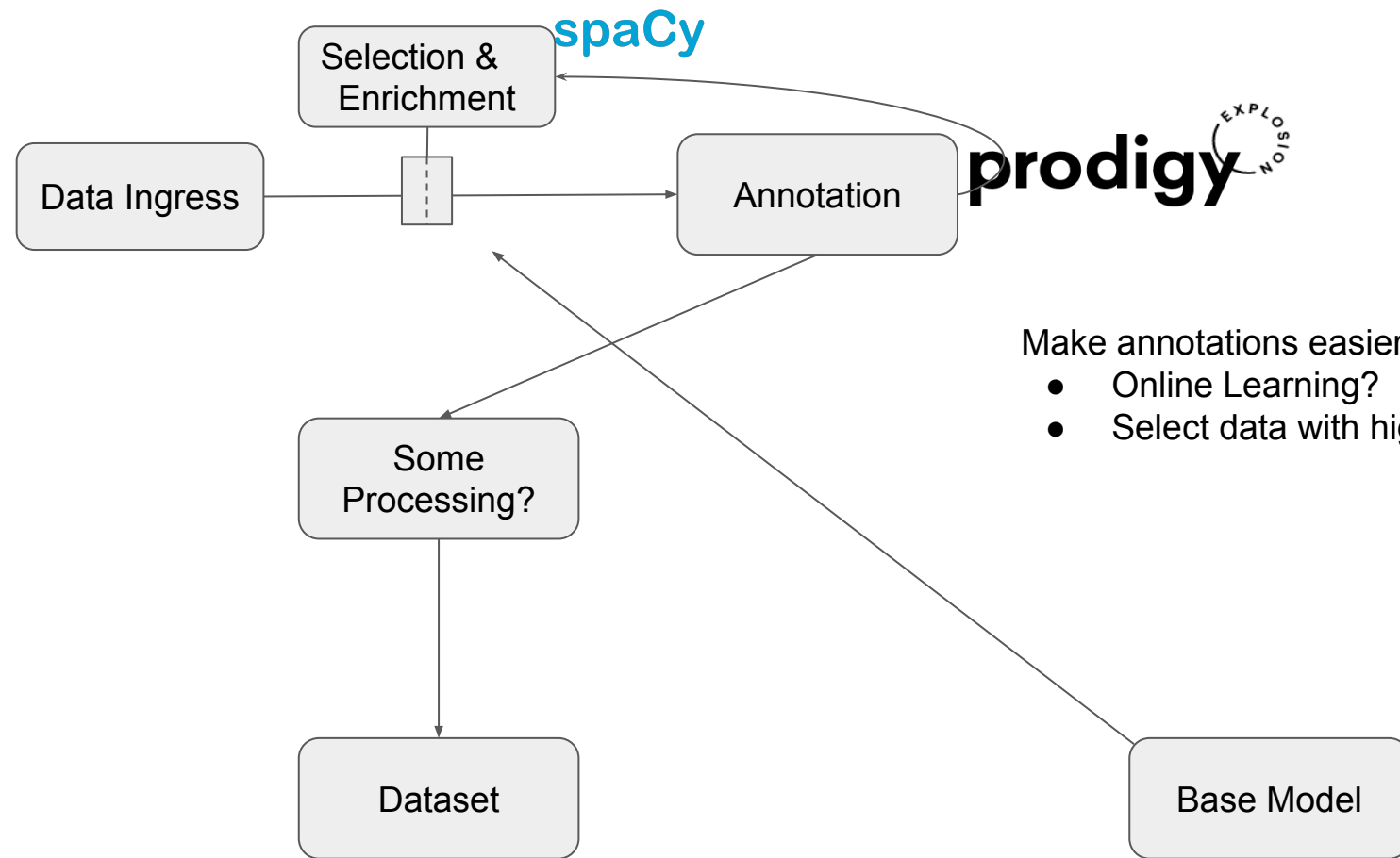


Dataset

The END ~ They trained happily ever after

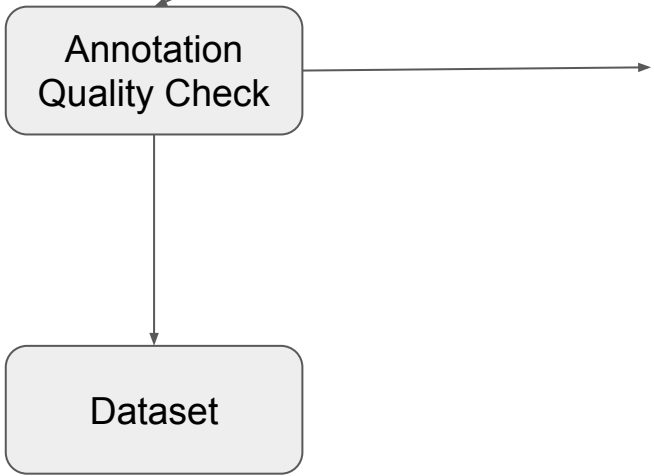
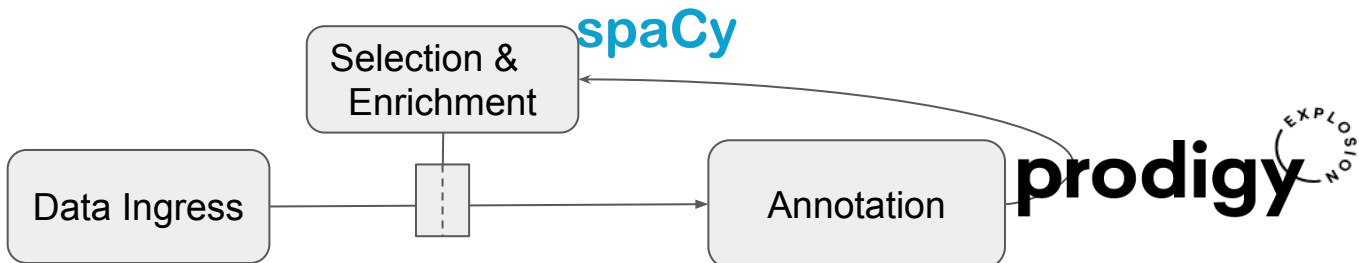


Make annotations easier to generate



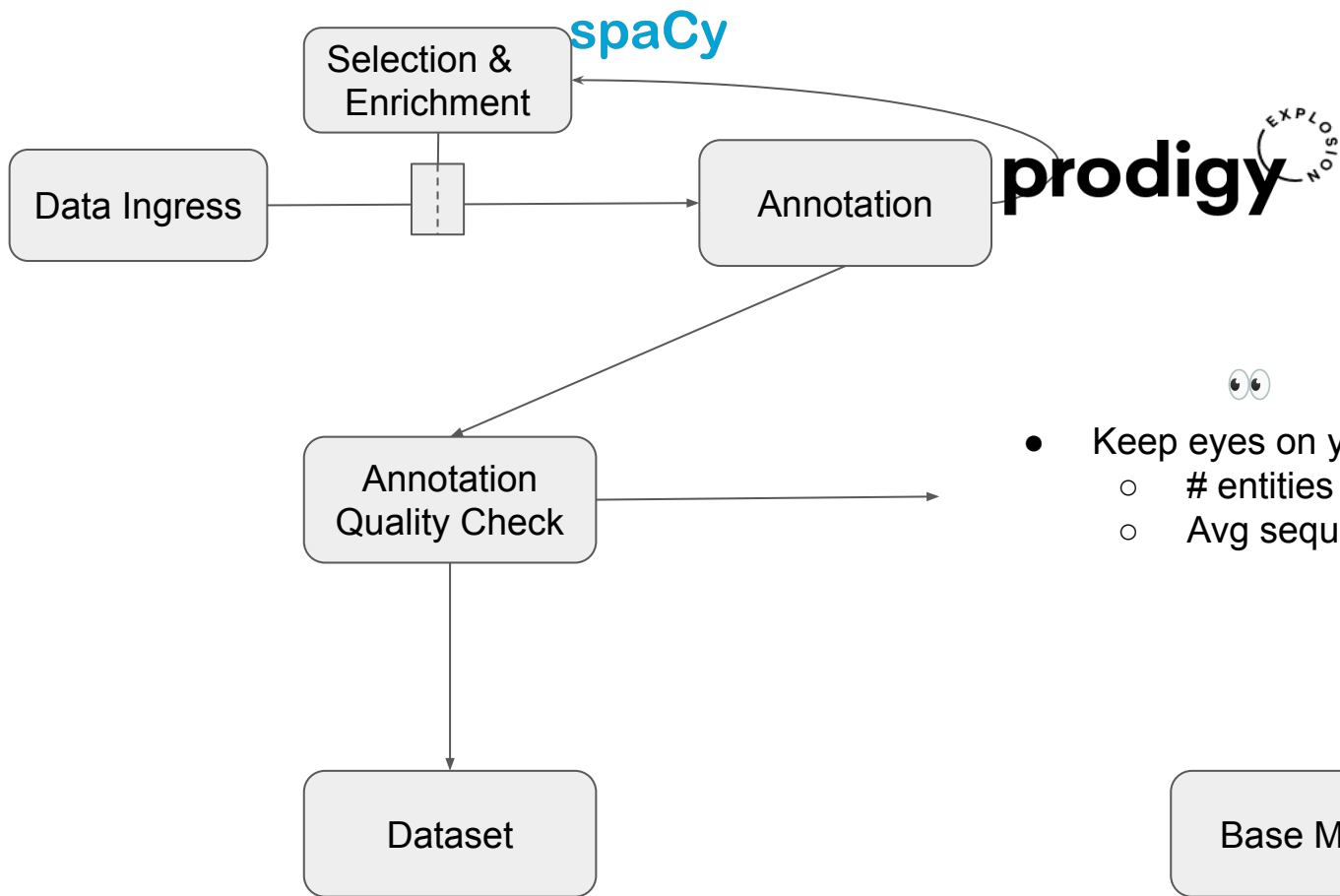
Make annotations easier to generate

- Online Learning?
- Select data with high information gain



- Keep eyes on your data stats
 - # entities per type / example
 - Avg sequence length



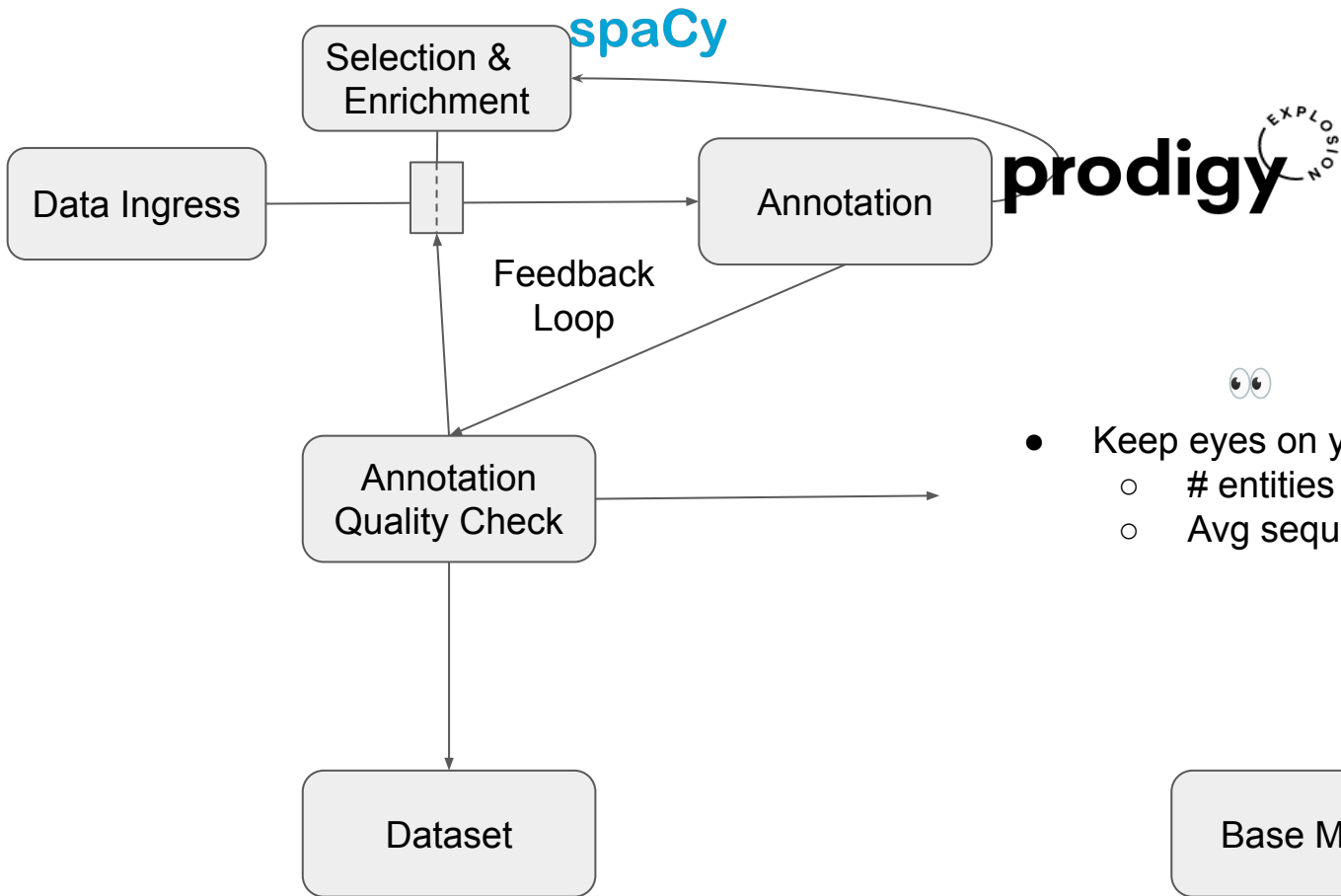


- Keep eyes on your data stats
 - # entities per type / example
 - Avg sequence length

```

NERStats(n_examples=87742,
n_examples_no_entities=135,
n_annotations=594811,
n_annotations_per_type=
{'ITEM': 318152, 'BRAND':
114962, 'COLOR': 88982,
'COMPOSITION': 72715},
examples_with_type=None)
  
```

Base M



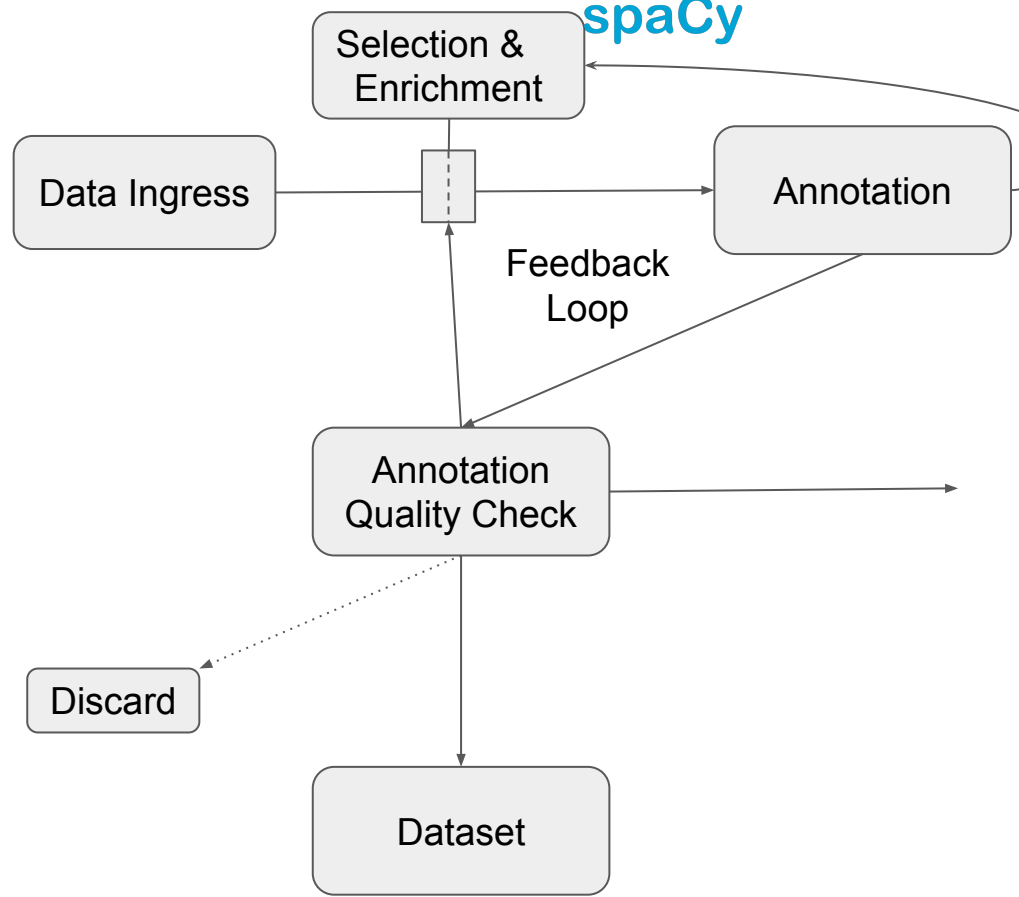
- Keep eyes on your data stats
 - # entities per type / example
 - Avg sequence length



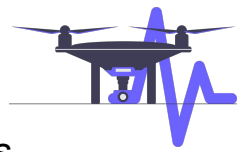
spaCy

EXPLOSION

prodigy

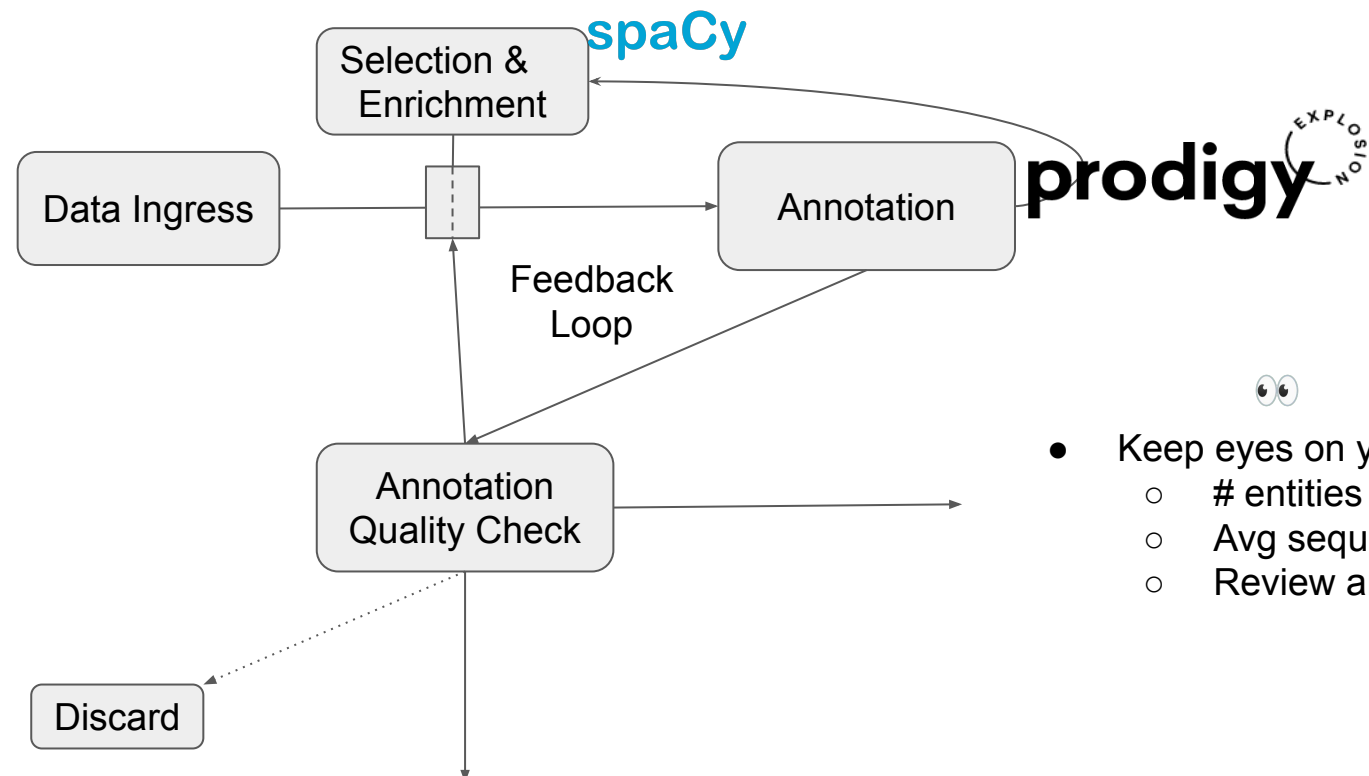


- Keep eyes on your data stats
 - # entities per type / example
 - Avg sequence length
 - Review annotations

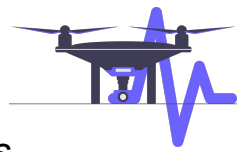


displaCy

Base Model



prodigy EXPLOSION



- Keep eyes on your data stats
 - # entities per type / example
 - Avg sequence length
 - Review annotations

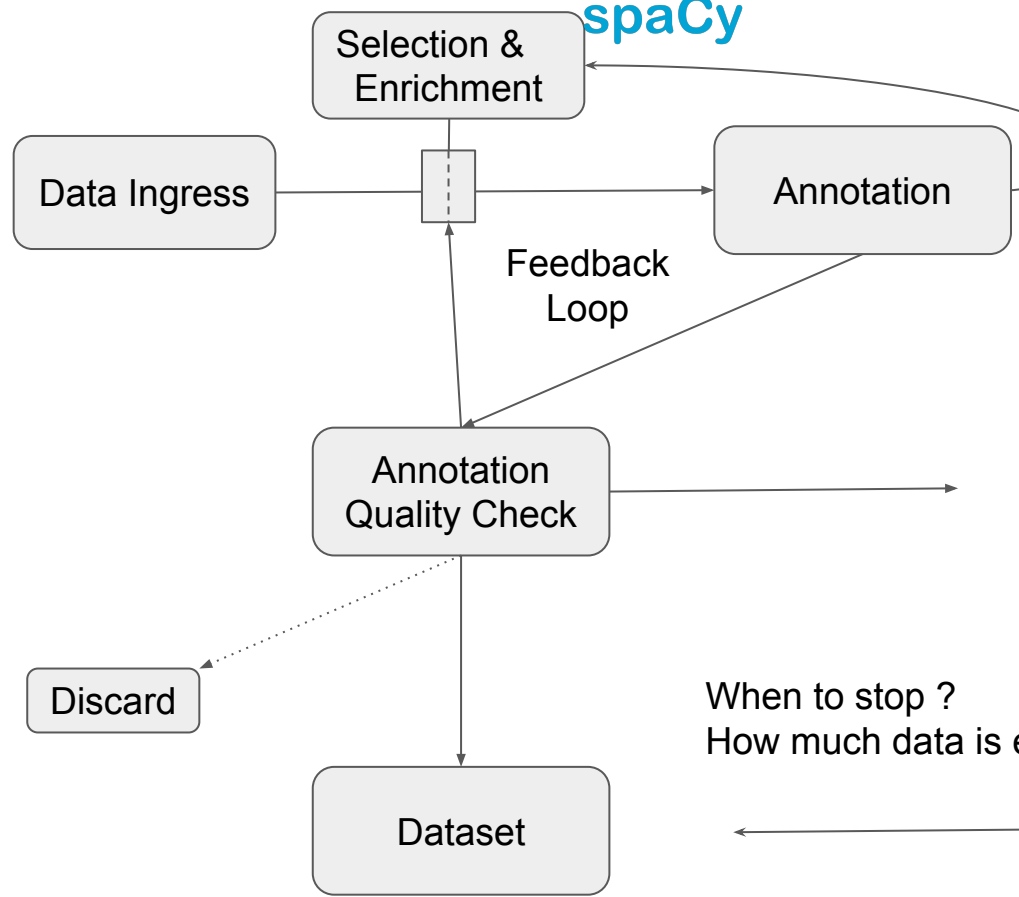
[displaCy](#)

Yellow COLOR ruffled wrap dress KEYWORD from Ganni BRAND featuring long sleeves KEYWORD and a wrap style front .

spaCy

EXPLOSION

prodigy



Feedback Loop



- Keep eyes on your data stats
 - # entities per type / example
 - Avg sequence length
 - Review annotations

displaCy

When to stop ?
How much data is enough?



prodigy ✓

PROJECT INFO
DATASET ner_tester

PROGRESS
THIS SESSION 0
TOTAL 18
0%

ACCEPT 0
REJECT 0
IGNORE 0

HISTORY

BRAND 1 COLOR 2 FABRIC 3 KEYWORD 4 OTHER 5

Blue color cotton blend faded slim jeans from
LIU JO featuring a stretch fit, a waistband with belt
loops, a button and zip fly, a five pocket
design and a regular length.



prodigy ✓

PROJECT INFO
DATASET ner_test_kr

PROGRESS
THIS SESSION 0
TOTAL 33
0%

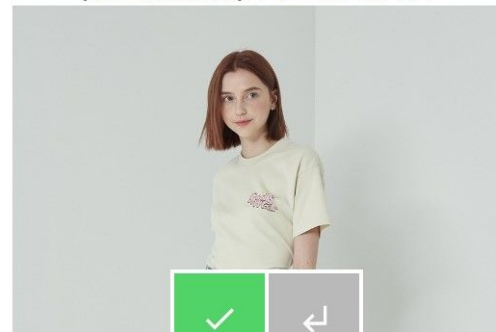
ACCEPT 0
REJECT 0
IGNORE 0

HISTORY

BRAND 1 COLOR 2 FABRIC 3 KEYWORD 4

MOTHER DENIM 은 미니멀한 감성을 담은 고급 캐주얼 웨어를 선보이는 브랜드입니다. 와이드 한 실루엣과 크롭 길이감의 데님 팬츠 입니다. 앞뒤 부분적으로 디스트로이드 디테일을 주었습니다. 100% 면 소재로 착용감이 좋습니다. 티셔츠 와 함께 매치하여 데일리한 캐주얼 룩을 완성할 수 있습니다. 프린트의 버튼 과 히든 버튼 으로 여미는 방식. 양 사이드와 힙에 총 5개의 포켓 . 피팅 정보 : 176 cm , 55 사이즈 모델이 61 사이즈 착용

[MOTHER DENIM] 와이드 크롭 데님 팬츠



prodigy ✓

PROJECT INFO
DATASET ner_tester

PROGRESS
THIS SESSION 0
TOTAL 18
0%

ACCEPT 0
REJECT 0
IGNORE 0

HISTORY

BRAND 1 COLOR 2 FABRIC 3 KEYWORD 4 OTHER 5

Blue color cotton fabric blend faded slim jeans keyword from
LIU JO brand featuring a stretch fabric fit, a waistband with belt
loops keyword, a button keyword and zip keyword fly, a five pocket
keyword design and a regular keyword length.



prodigy ✓

PROJECT INFO
DATASET ner_test_kr

PROGRESS
THIS SESSION 0
TOTAL 33
0%

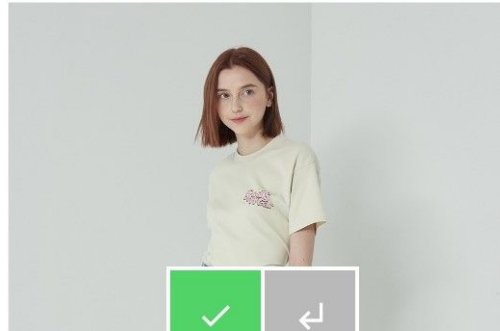
ACCEPT 0
REJECT 0
IGNORE 0

HISTORY

BRAND 1 COLOR 2 FABRIC 3 KEYWORD 4

MOTHER DENIM brand 은 미니멀한 감성을 담은 고급 캐주얼 keyword
웨어를 선보이는 브랜드입니다. 와이드 keyword 한 실루엣과 크롭
keyword 길이감의 데님 keyword 팬츠 keyword 입니다. 앞뒤 부분적으
로 디스트로이드 keyword 디테일을 주었습니다. 100% 면 fabric 소
재로 착용감이 좋습니다. 티셔츠 keyword 와 함께 매치하여 데일리한
캐주얼 keyword 룩을 완성할 수 있습니다. 프린트의 버튼 keyword 과
히든 버튼 keyword 으로 여미는 방식. 양 사이드와 힙에 총 5 개의 포
켓 keyword . 피팅 정보 : 176 cm , 55 사이즈 모델이 61 사이즈 착용

[MOTHER DENIM] 와이드 크롭 데님 팬츠



Couldn't use pattern matching pre-annotation enrichment

...면(entity) ...면(O)

Overview

- Problem Definition
- Data
 - Collection
 - Annotation
 - Processing
- Modelling
 - Train
 - Evaluate
- Deploy
 - Server
 - UI



Overview

- Problem Definition
- Data
 - Collection
 - Annotation
 - Processing
- Modelling
 - Train
 - Evaluate
- Deploy
 - Server
 - UI

%85



Overview

- Problem Definition
- Data
 - Collection
 - Annotation
 - Processing
- Modelling
 - Train
 - Evaluate
- Deploy
 - Server
 - UI

%85



%15 or less if you use the right tools!

The Easy Part



HUGGING FACE



Weights & Biases

The Easy Part

Frameworks



HUGGING FACE

+




Weights & Biases

Models







BERT (eng)
ELECTRA (kr)

The Easy Part: Just Plug your data

🔑 master ▾ transformers / examples / pytorch / token-classification / Go to file Add file ▾ ...

 **sgugger** Add option to log only once in multinode training (#11819) ... ✖ f086652 11 days ago 🕒 History

..

 README.md	Add more subsections to main doc (#11758)	18 days ago
 requirements.txt	Examples reorg (#11350)	2 months ago
 run.sh	Examples reorg (#11350)	2 months ago
 run_ner.py	Add option to log only once in multinode training (#11819)	11 days ago
 run_ner_no_trainer.py	Examples reorg (#11350)	2 months ago
 run_no_trainer.sh	Examples reorg (#11350)	2 months ago

☰ README.md

Token classification

Tokenization - Wordpiece Basic Approach

cotton	pleated ...	philipp plein
B-Col	B-Key ...	B-Brand I-Brand

cotton	plea ##ted	p ##hil ##ip ##p	p ##lein
B-Col	B-Key X	B-Brand X X	I-Brand X

##subword's

X-> ignored during Training



Tokenization - Wordpiece Basic Approach

cotton	pleated ...	philipp plein
B-Col	B-Key ...	B-Brand I-Brand

cotton	plea ##ted	p ##hil ##ip ##p	p ##lein
B-Col	B-Key X	B-Brand X X	I-Brand X

##subword's

X-> ignored during Training



Ner Pipeline: Entity grouping 🤗

[WIP] Ner pipeline grouped_entities fixes #5970

Merged LysandreJik merged 23 commits into `huggingface:master` from `cceyda:ner_pipeline_group_fix` on Nov 4, 2020

Conversation 51 Commits 23 Checks 0 Files changed 2



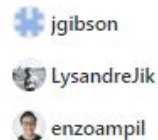
cceyda commented on Jul 22, 2020 • edited

Contributor 🗨️ ⋮

There are many issues with ner pipeline using `grouped_entities=True`

#5077
#4816
#5730
#5609
#6514
#5541

Reviewers



Assignees

No one assigned

🎉 Bug fix contribution



Subword Aggregation strategies

Instead of ignoring subwords use them!

[TokenClassification] Label realignment for subword aggregation #11680

Merged Narsil merged 6 commits into `huggingface:master` from `Narsil:token_classification_aggregation_strategy` 18 days ago

Conversation 39 Commits 6 Checks 1 Files changed 4



Narsil commented 25 days ago • edited

Contributor

What does this PR do?

Tentative to replace #11622

- Added `AggregationStrategy`

- Reviewers
- francescorubbo
 - cceyda
 - LysandreJik

Assignees

↑ Proposal: Non-NER span categorizer #3961

1

honnibal started this conversation in New Features & Project Ideas



honnibal on Jul 13, 2019 Maintainer

edited ...

People often want to train spaCy's `EntityRecognizer` component on sequence labeling tasks that aren't "traditional" Named Entity Recognition (NER). A named entity actually has a fairly specific linguistic definition: they're either definite descriptors, or a numeric expression. Software designed for doing NER may not perform well on other sequence labelling tasks. This is especially true of spaCy's NER system, which makes fairly aggressive trade-offs. What we need is another system that's better suited for these non-NER sequence tasks.

Add SpanCategorizer component #6747

Draft

honnibal wants to merge 41 commits into `master` from `feature/span-categorizer-v3`

Conversation 35

Commits 41

Checks 8

Files changed 11



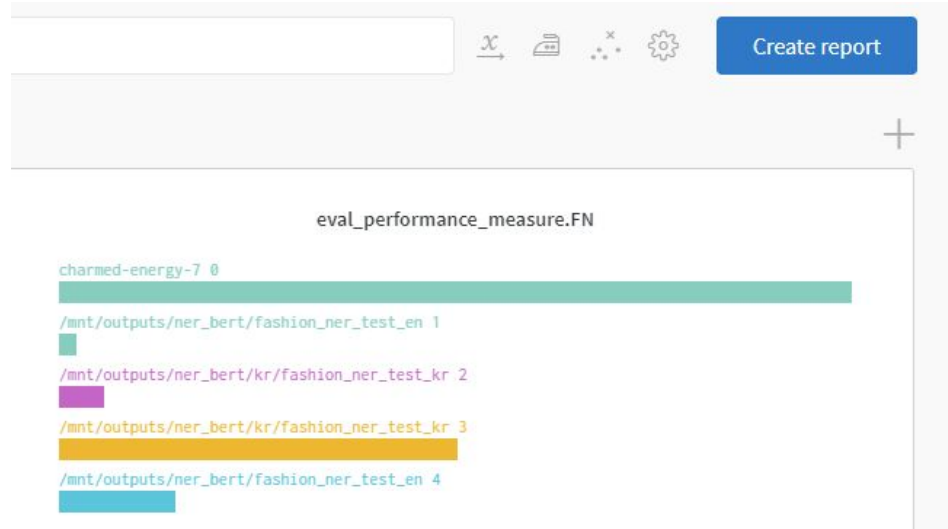
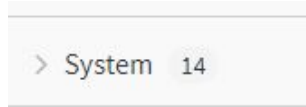
honnibal commented on Jan 17 • edited

Member

Experiment Tracking with Weights & Biases

Basic Metrics;

- Precision, Recall, F1
- Resource Usage / Time



Add additional Metrics

- Sentence/word/char avg scores
- Confusion Matrix (used sklearn 😊)



Model Deployment: Demos & Beyond





Streamlit 🤖 Build Web Apps with Python

FashionNER

Input

silk blue mini skirt by Designovel

silk FABRIC

blue COLOR

mini skirt KEYWORD

by

Designovel BRAND

Korean FashionNER

Input

레오파드 패턴 치마

레오파드 FABRIC

패턴

치마 KEYWORD



Streamlit



Build Web Apps with Python

```
1 from spacy import displacy
2 import os
3 import streamlit as st
4 from utils import hf_ents_to_displacy_format, make_color_palette
5 import httpx
6
7 HTML_WRAPPER = """<div style="overflow-x: auto; border: 1px solid #e6e9ef; border-radius: 0.25rem; padding:
8 1rem; margin-bottom: 2.5rem">{</div>"""
9
10 @st.cache()
11 def load_model():
12     labels = ["B-BRAND", "B-COLOR", "B-FABRIC", "B-KEYWORD", "I-BRAND", "I-COLOR", "I-FABRIC", "I-KEYWORD", "O"]
13     color_map = make_color_palette([l.split("-")[-1] for l in labels])
14     return color_map
15
16 def predict(input_):
17     # confirm input str
18     res = httpx.post("http://127.0.0.1:7863/predictions/fashioner", data=input_)
19     return res.json()
20
21 def display(bert_ents):
22     bert_doc = hf_ents_to_displacy_format(bert_ents, ignore_entities=["O"])
23     html = displacy.render(bert_doc, manual=True, style="ent", options={"colors": color_map})
24
25     html = html.replace("\n", " ") # Newlines seem to mess with the rendering
26     st.write(HTML_WRAPPER.format(html), unsafe_allow_html=True)
27
28 color_map = load_model()
29
30 st.header("FashionER")
31 input_ = st.text_input("Input", "silk blue mini skirt by Designovel")
32 bert_ents = predict(input_)
33 display(bert_ents)
```

Just 30~50 lines

FashionER

Input

silk blue mini skirt by Designovel

silk FABRIC

blue COLOR

mini skirt KEYWORD

by

Designovel BRAND

Korean FashionER

Input

레오파드 패턴 치마

레오파드 FABRIC

패턴

치마 KEYWORD

Additional Steps & Future Work

You are now almost production ready~

- API security & monitoring
- Model Performance & Statistic monitoring

Finer Grained NER

- more entity types
- nested types

Summary & Advice

- Use Open Source already available
- Contribute
- Play with different datasets
- Create your own datasets
- Don't be afraid to try new things



@cceyda



@ceyda_cinarel

Thank You 