John Snow LABS

Delivering Successful Data Annotation Projects

Dia Trambitas, PhD, Head of Product

Andrei Feier, MD, PhD, Lead Clinical Annotator

Radu Bisca, NLP Lab Product Manager

Outline

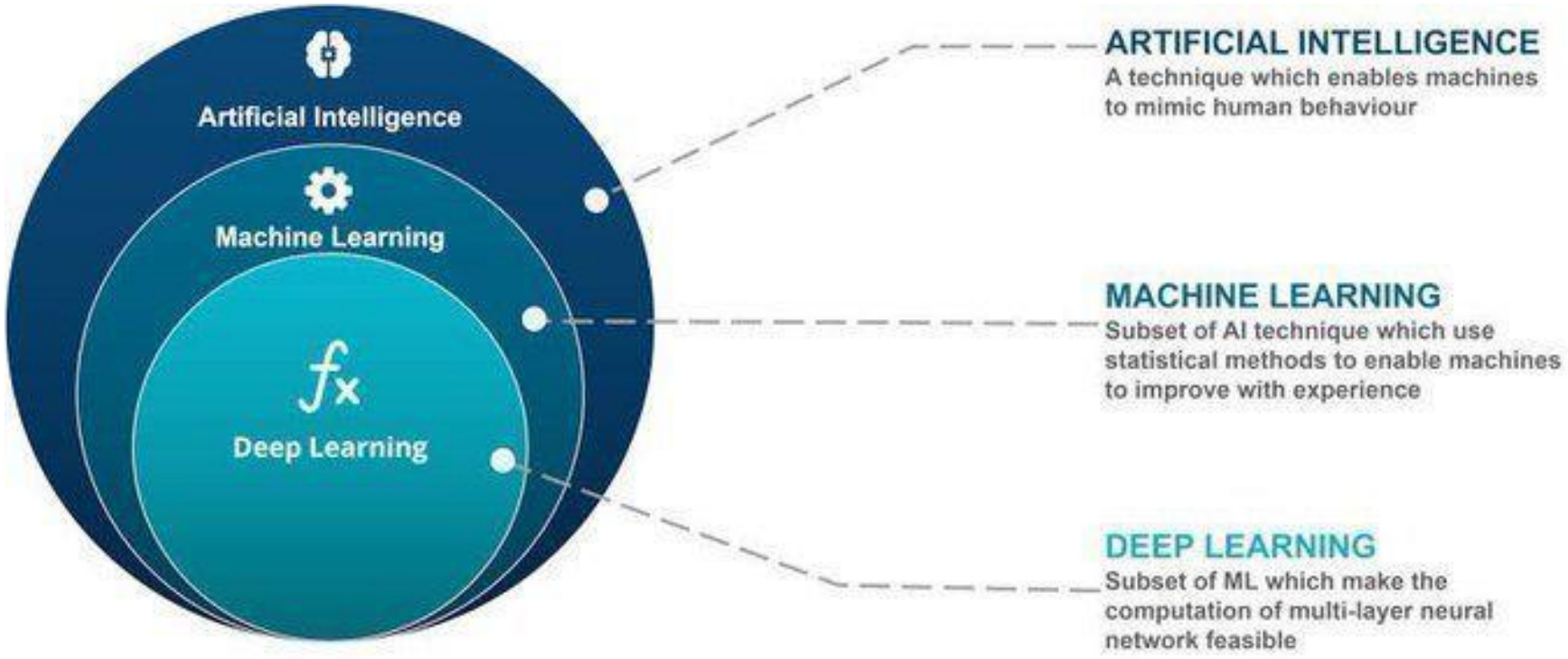
1. Introduction to Text Annotation
2. Annotation Projects Setup in NLP Lab
3. Annotation Guidelines
4. Preannotation Resources
5. Model Training
6. Conclusions and further resources

Outline

Context

John Snow LABS

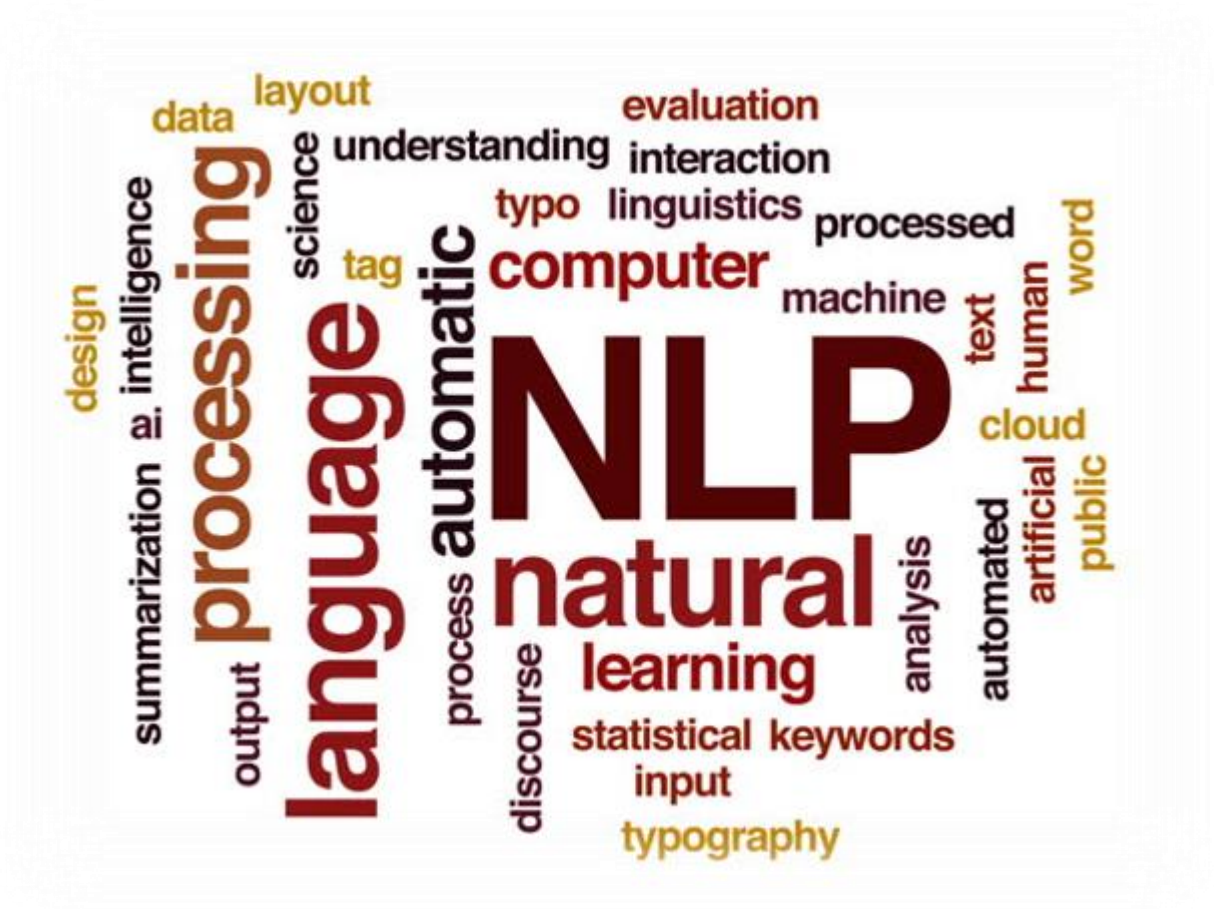
- ML is the study of computer algorithms that improve automatically through experience.
- ML models make decisions based on sample data (training data), without being explicitly programmed.



Natural Language Processing

John Snow LABS

- NLP is a subfield of linguistics and computer science concerned with the processing and analysis of natural language data.

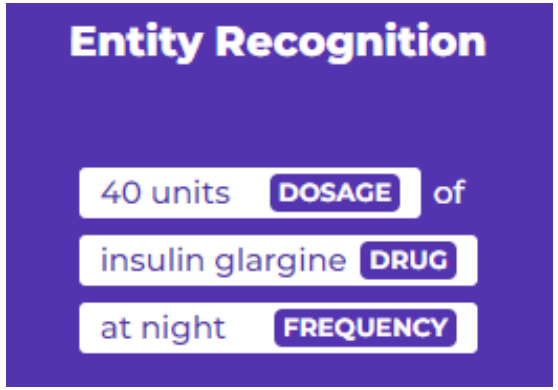


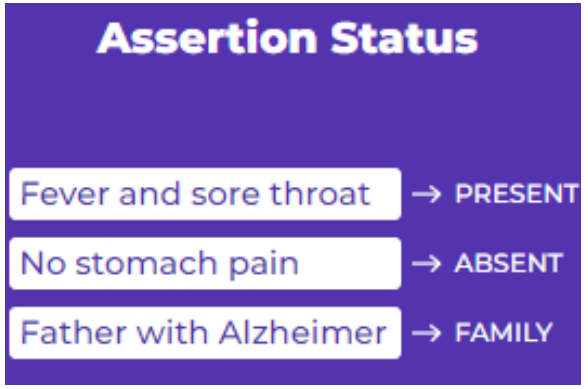
What is Data Annotation?

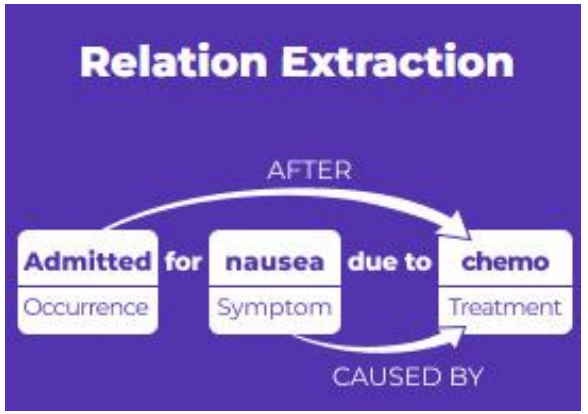
- The process of labeling or tagging raw data with meaningful information, making it usable for training machine learning models.

Annotation Tasks in NLP

John Snow LABS







Why is Data Annotation Crucial for NLP?

- Transforming unstructured data into **structured** format.
- Providing "ground truth" for model training and validation.
- Enhancing the performance of NLP models.
- Ensuring models understand context, sentiment, entities, etc.
- A way of transferring annotators knowledge and experience to models.

- **The resulting models can only be as good as the training data.**

Role of Annotation Tools

- Facilitate **systematic and organized annotation**.
- Allow for **collaborative annotation** projects.
- Ensure standardization and **consistency**.
- Offer **quality control** and **review** functionalities.
- Provide **automation** and **AI-assisted features**.

Key Requirements for Annotation Tools

- **High productivity** – keep annotators "in the zone"
  - ✓ Intuitive, easy to learn
  - ✓ Keyboard shortcuts
  - ✓ How many clicks? How many eye movements?
  - ✓ Automatic transition between documents
- **Teamwork**
  - ✓ Support for projects, teams, and role-based permissions
  - ✓ Customizable workflows: validation, review, model training, etc.
  - ✓ Collaboration: compare versions, share guidelines
  - ✓ Versioning & audit trails
- **Improve over time**
  - ✓ New features and constant improvements
  - ✓ Learn from annotators actions
  - ✓ High-accuracy, configurable deep learning backend
  - ✓ "Closed-loop" automation: pre-train, annotate, retrain, measure

Let’s get started!

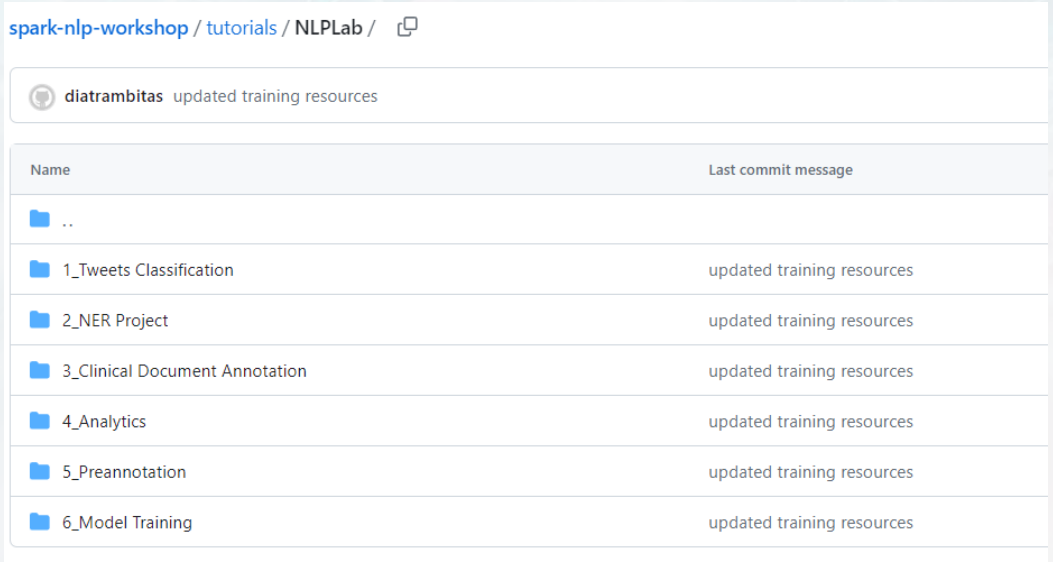
Training Resources

Dedicated **NLP Lab Instance** - http://3.216.238.207/

- m4.4xlarge EC2 instance – 16 CPUs and 64GB RAM
- available for 2 weeks
- **Credentials shared via Email** (check also in the spam folder)

Git Hub - https://github.com/JohnSnowLabs/spark-nlp-workshop

- Clone the repo on your local machine
- Resources available under **/spark-nlp-workshop/tutorials/NLPLab**



Outline

1. Introduction to Text Annotation
2. **Annotation Projects Setup in NLP Lab**
3. Annotation Guidelines
4. Preannotation Resources
5. Model Training
6. Conclusions and further resources

Project Setup

1. Project creation steps for Classification project
   - Configuration, choices, hotkeys
2. Project creation steps for NER project
   - Configuration, choices, hotkeys
3. Task import
4. Task assignment
5. Workflow details

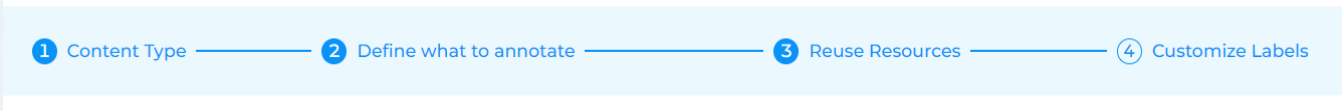
Project Creation/Classification

1. Login into NLP Lab
2. Once logged in, in the projects page, select "New Project"

3. In the Project Description page, choose a name for your project and click Next. The name must be unique, you will get a warning in case there is already a project with the name you choose.
4. In the Add team Members page, use the "Search team Members" box to lookup usernames you want added to the project

   Note: the users must be created and added to the system prior to assigning them to a project

   Click Next. This takes you to the Configuration page. Configuring your project is a guided experience and has 4 steps displayed in the top horizontal bar:

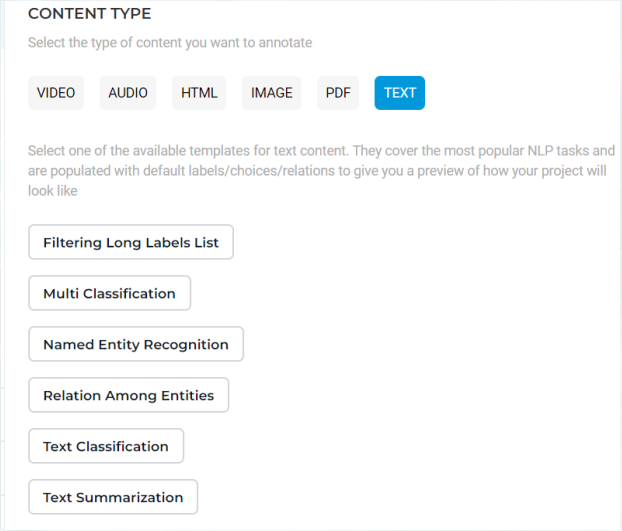


   Select Text Classification and click Next.

Project Creation/Classification

For the training session today, we will create 2 different project types : Text Classification and Named Entity Recognition

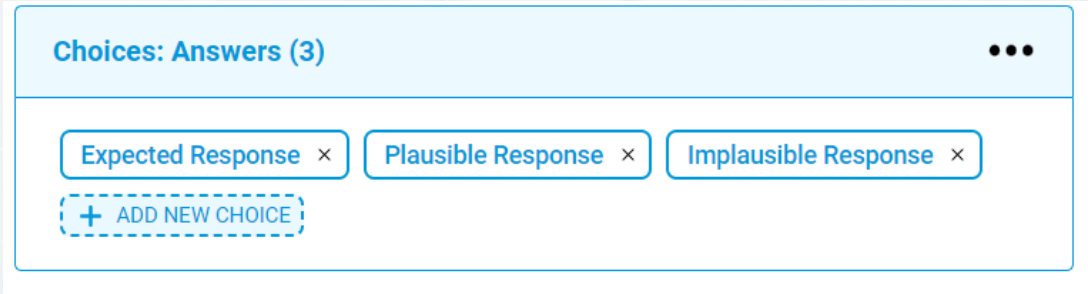
5. Select Text Classification and click Next.



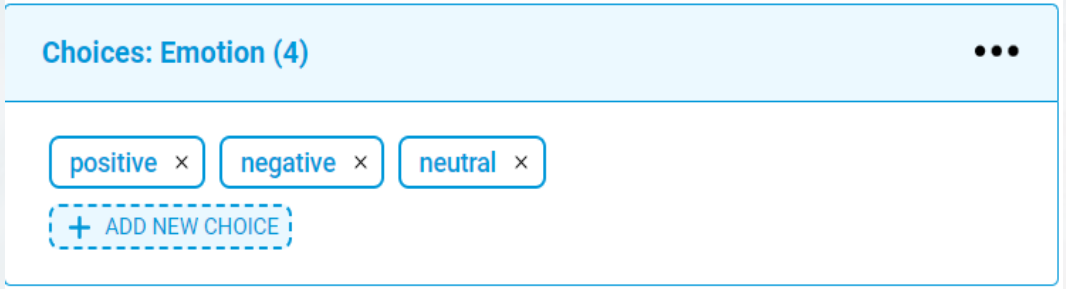
6. In Step 2 and 3 leave the default settings and click Next.

Project Creation/Classification-cont

7. In Step 4 "Customize Labels“, click on Choices.
By default, few choices are already populated for you.



Add yours then remove existing ones then click the Save Config button.
We will be classifying tweets in this project so the choices we add are “positive”, “negative” and “neutral”



This takes us to the Import page where tasks will be imported into the project.

Project Creation/Classification-cont

- Customize how you view your choices

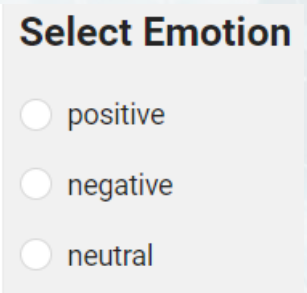
By default, the choices are listed vertically. If you prefer to list them horizontally, switch to Code view and add “showInLine” tag set to True

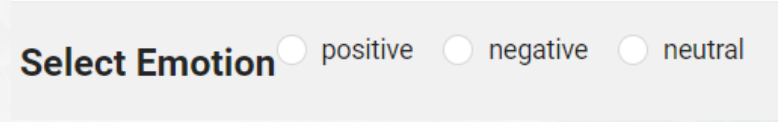
CUSTOMIZE CONFIGURATION

Visual | Code

Select from following custom options category

```
1 View orientation="horizontal" oneClickSubmit="true">
2  <Choices name="Emotion" toName="text" choice="single" showinline = "true"
3    <Header value="Select Emotion"/>
4    <Choice value="positive"/>
5    <Choice value="negative"/>
6    <Choice value="neutral"/>
7  </Choices>
8  <Text name="text" value="$text"/>
9 /View>
10
```





Project Creation/NER

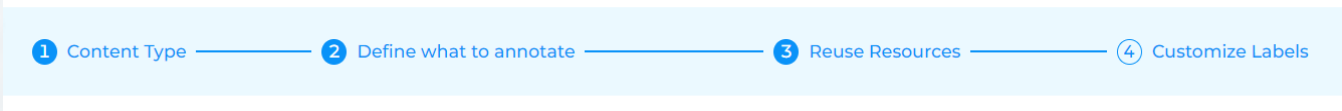
1. Login into NLP Lab
2. Once logged in,
in the projects page, select "New Project"

3. In the Project Description page, choose a name for your project and click Next. The name must be unique, you will get a warning in case there is already a project with the name you choose.

4. In the Add team Members page, use the "Search team Members" box to lookup usernames you want added to the project

Note: the users must be created and added to the system prior to assigning them to a project

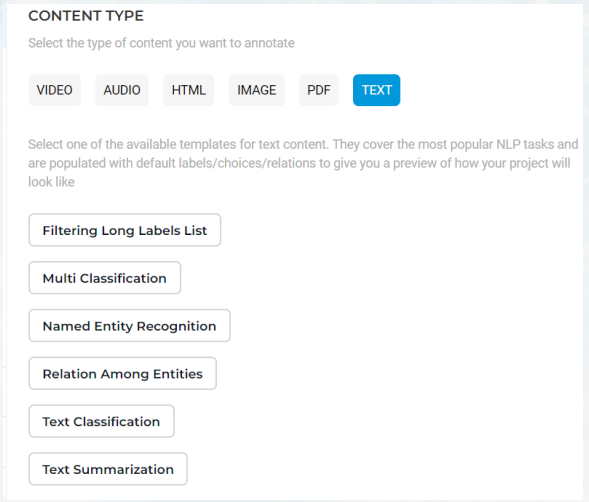
Click Next. This takes you to the Configuration page. Configuring your project is a guided experience and has 4 steps displayed in the top horizontal bar:



Project Creation/NER

For the training session today, we will create 2 different project types : Text Classification and Named Entity Recognition

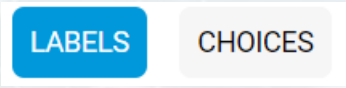
5. Select Named Entity Recognition and click Next.



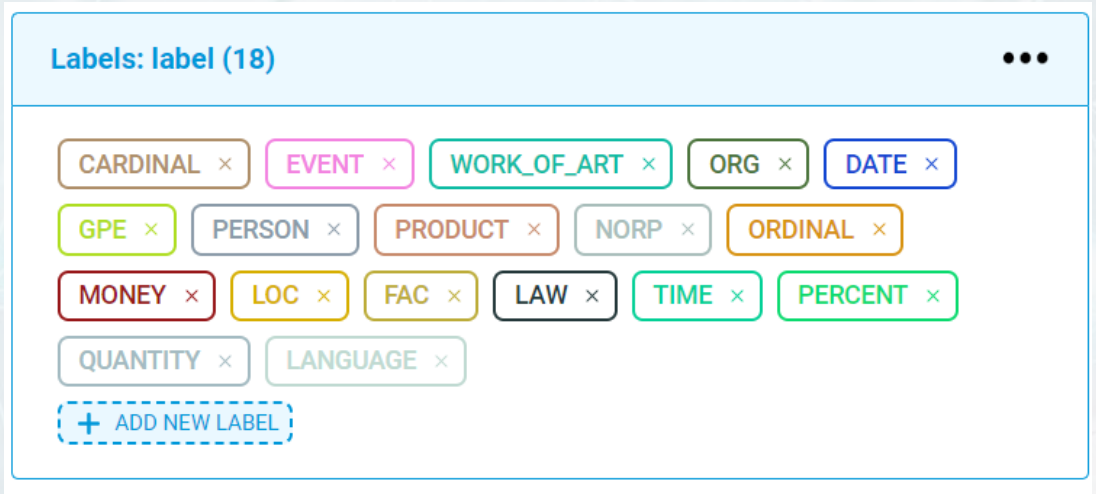
6. In Step 2 and 3 leave the default settings and click Next.

Project Creation/NER-cont

7. In Step 4 "Customize Labels“, click on Labels.



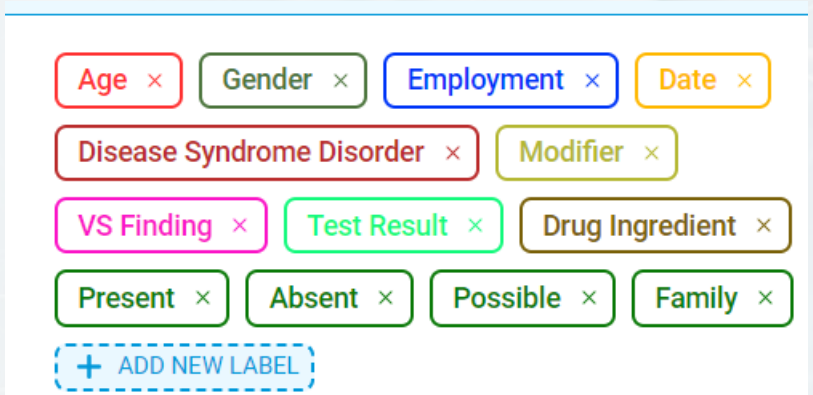
By default, few labels are already populated for you.



Add yours then remove existing ones then click the Save Config button. This takes us to the Import page where tasks will be imported into the project.

Project Creation/NER-cont

- Tip: Remove unwanted labels by editing the xml config file. You can do this by switching modes between Visual to Code




```
<View orientation="horizontal">
  <Labels name="label" toName="text">
    <Label value="Age" background="#ff3333"/>
    <Label value="Gender" background="#4c753d"/>
    <Label value="Employment" background="#0036f8"/>
    <Label value="Date" background="#ffb800"/>
    <Label value="Disease Syndrome Disorder" background="#ba2b2b"/>
    <Label value="Modifier" background="#b4b731"/>
    <Label value="VS Finding" background="#fb16c8"/>
    <Label value="Test Result" background="#13fa7c"/>
    <Label value="Drug Ingredient" background="#7a6109"/>
    <Label value="Present" background="#0b7a09" assertion="true"/>
    <Label value="Absent" background="#0b7a09" assertion="true"/>
    <Label value="Possible" background="#0b7a09" assertion="true"/>
    <Label value="Family" background="#0b7a09" assertion="true"/>
  </Labels>
  <Text name="text" value="$text"/>
</View>
```

Project Creation/NER-cont

- Creating hot keys.

Hot keys are keyboard shortcuts you can set for increasing productivity in the labeling process.
To setup hot keys, switch to Code mode and add hotkey tag:

```
<View orientation="horizontal">
  <Labels name="label" toName="text">
    <Label value="Age" background="#ff3333" hotkey="A"/>
    <Label value="Gender" background="#4c753d" hotkey="G"/>
    <Label value="Employment" background="#0036f8" hotkey="E"/>
    <Label value="Date" background="#ffb800"/>
    <Label value="Disease Syndrome Disorder" background="#ba2b2b"/>
    <Label value="Modifier" background="#b4b731"/>
    <Label value="VS Finding" background="#fb16c8"/>
    <Label value="Test Result" background="#13fa7c"/>
    <Label value="Drug Ingredient" background="#7a6109"/>
    <Label value="Present" background="#0b7a09" assertion="true"/>
    <Label value="Absent" background="#0b7a09" assertion="true"/>
    <Label value="Possible" background="#0b7a09" assertion="true"/>
    <Label value="Family" background="#0b7a09" assertion="true"/>
  </Labels>
  <Text name="text" value="$text"/>
</View>
```

Project Creation/NER-cont

- Assertion Labels

Assertion labels are tags with contextual information that are assigned to NER extractions. The typical example of assertion status detection is negation identification: in the sentence “the patient has no history of diabetes”, the word “diabetes” is annotated using both an NER label (Disease) and an assertion label (Absent).

These labels are defined using the configuration file as a “regular” label + assertion=“true”

```
    <Label value="Present" background="#0b7a09" assertion="true"/>
    <Label value="Absent" background="#0b7a09" assertion="true"/>
    <Label value="Possible" background="#0b7a09" assertion="true"/>
    <Label value="Family" background="#0b7a09" assertion="true"/>
```

Using the Annotation Lab, you can annotate assertion in a very simple way. First, choose a NER label and select the part of the text that you want to extract, and then choose an assertion label and select that same part of the document.

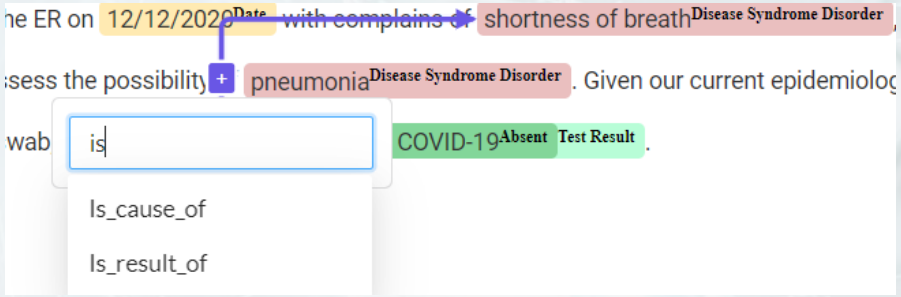
Read more about assertion labels here - https://www.johnsnowlabs.com/tips-and-tricks-on-how-to-annotate-assertion-in-clinical-texts/

Project Creation/NER-cont

- Creating relations

Relations can be defined using the configuration file by adding a Relation section as depicted in the screenshot below.

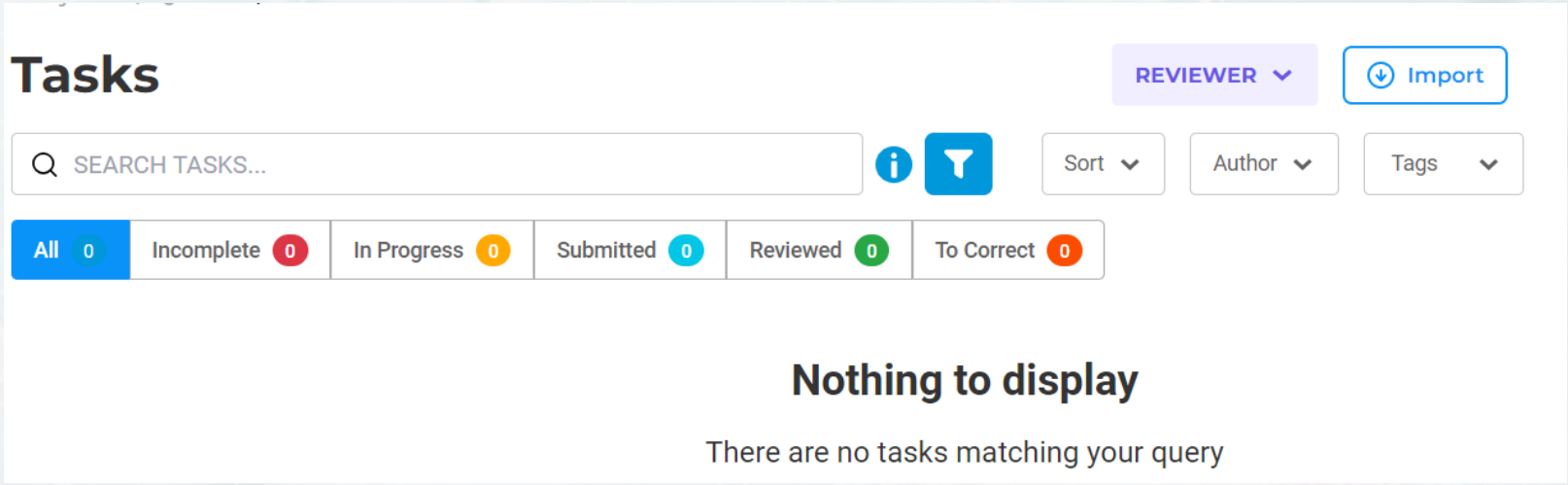
```
  </Labels>
  <Relations>
        <Relation value="Is_cause_of"/>
        <Relation value="Is_result_of"/>
</Relations>
```



When creating the relation, the name defined in the configuration file is displayed in the UI.

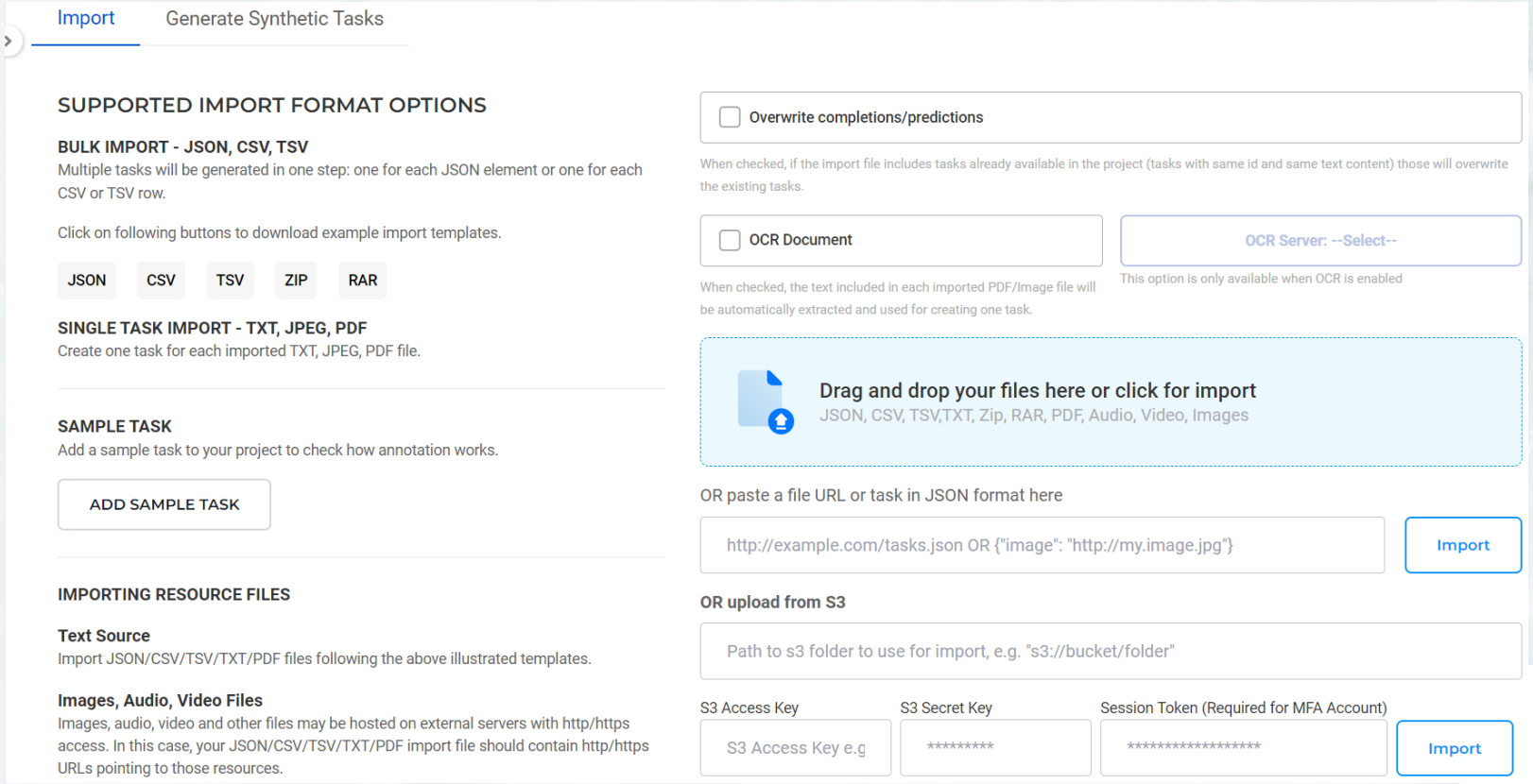
Task Import and Assignment

When project setup is done, you will land on the Task page. The page is empty, and to import your tasks, click on the Import button – top right.



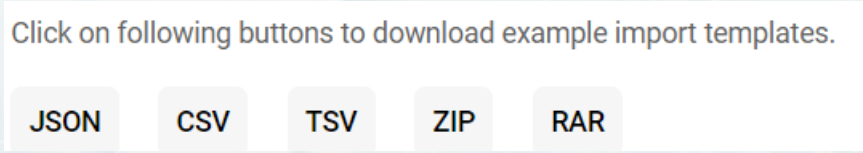
Task Import and Assignment

There are multiple ways for importing tasks,: using a zip archive with your text files, importing tasks using a csv file, using json file, etc



Task Import and Assignment

Download any sample file by clicking on the specific icon.

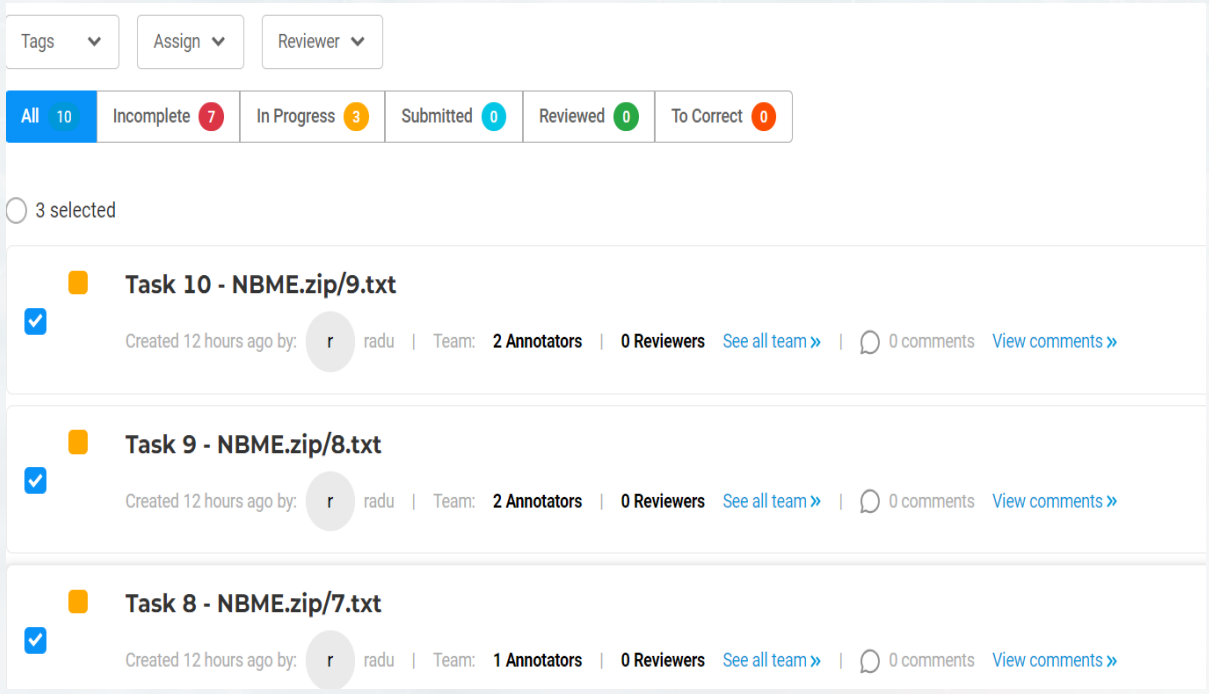


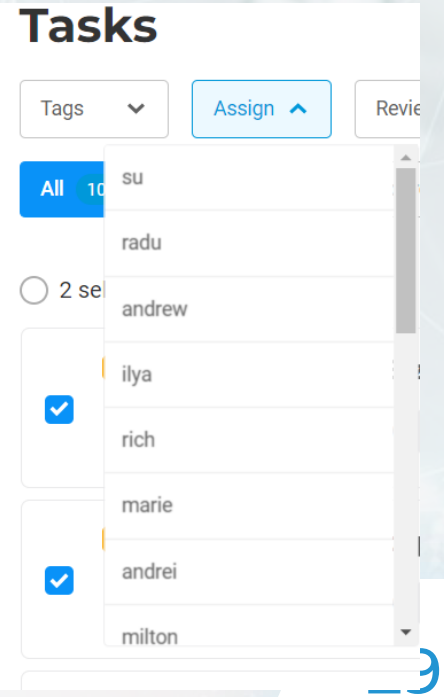
The screenshot below shoes a sample file used to import tweets in the classification project used as part of this training sessionCSV file used for this demo:

text	title
I`d have responded, if I were going	t1
Sooo SAD I will miss you here in San Diego!!!	t2
my boss is bullying me...	t3
what interview! leave me alone	t4
Sons of ****, why couldn`t they put them on the releases we already bought	t5

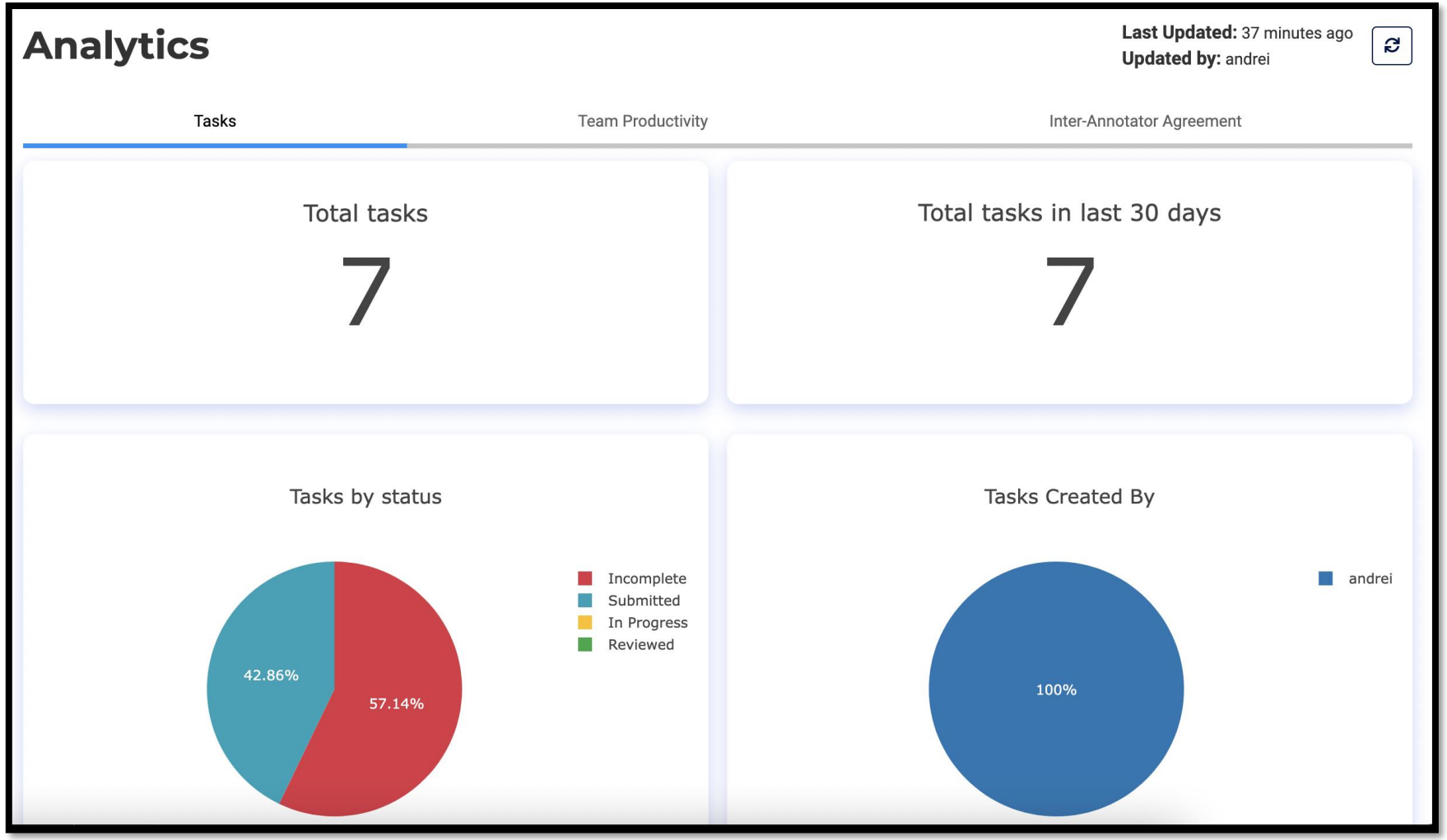
Task Import and Assignment

Once you import the tasks, choose to Explore tasks; this takes you to the Tasks page. As a project owner you will start assigning tasks to Annotators.
select one or more tasks, click on Assign drop down and select the name of the user to assign selected tasks. Repeat this step to continue assigning tasks.





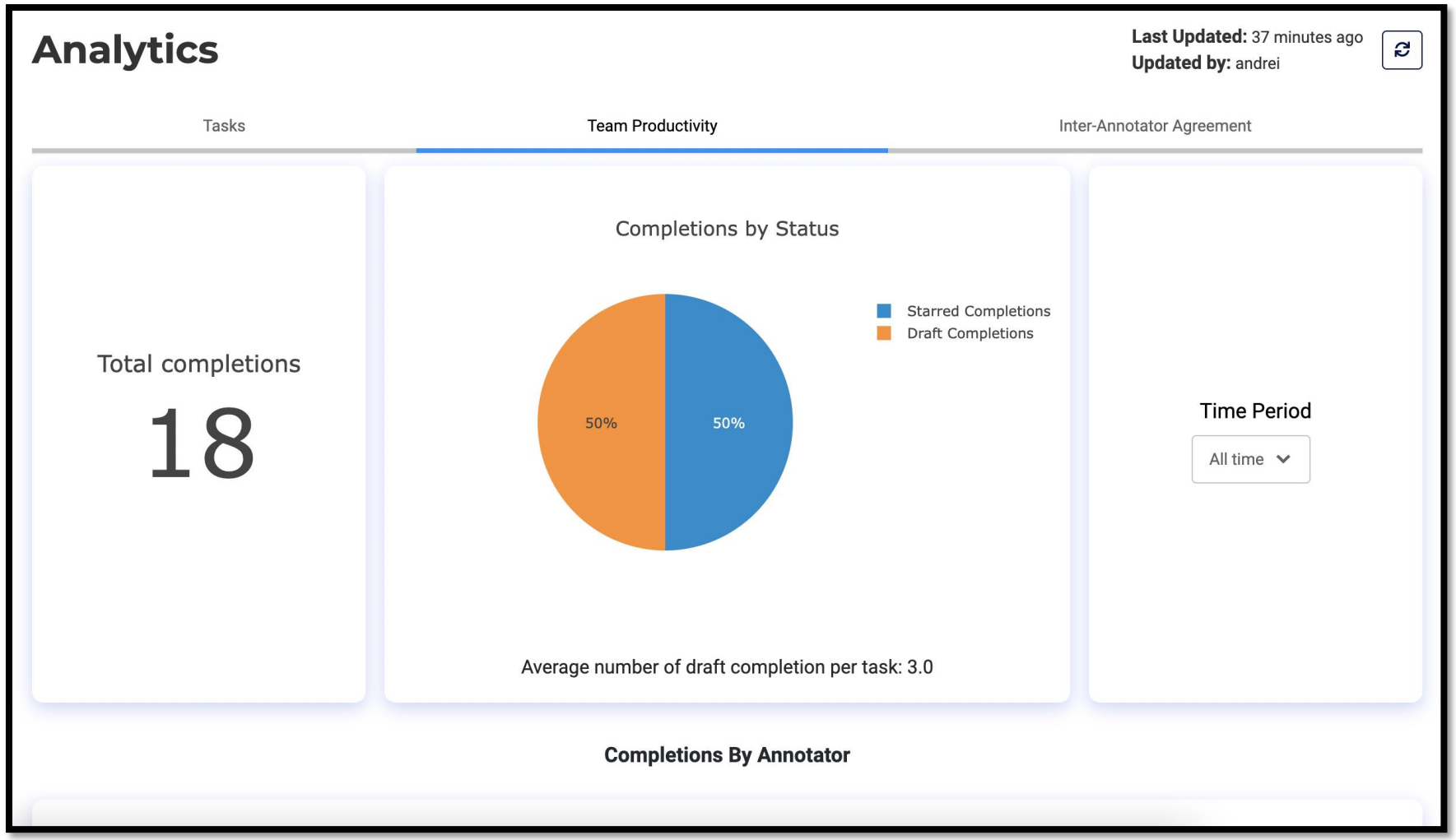
Analytics of NLP Lab - Tasks



- different charts
- overall annotation progress

Click here for in-depth explanation

Analytics of NLP Lab - Productivity



- different tools
- productivity across team

Click here for in-depth explanation

Analytics of NLP Lab – IAA

- level of agreement
- tools for assessing IAA

Click here for in-depth explanation

Outline

1. Introduction to Text Annotation
2. Annotation Projects Setup in NLP Lab
3. **Annotation Guidelines**
4. Preannotation Resources
5. Model Training
6. Conclusions and further resources

John Snow LABS

Annotation Guidelines (AG)

Andrei Feier, MD, PhD

Clinical Lead Annotator

Contents

John Snow LABS

Stakeholders and Roles

Stakeholders and Roles

John Snow LABS

Annotator Lead

- oversees the entire project - > guidelines align with the project's objectives
- keeps track of changes with a Change Log
- manager role in projects

Annotators

- usually medical professionals (MDs, nurses etc.)
- assess/curate data/annotate
- annotator/reviewer role in projects

Stakeholders and Roles

John Snow LABS

Data Scientist

- technical stakeholders who will use the annotated data for training
- provide requirements for annotation granularity and specificity

Quality Assurance (Reviewers)

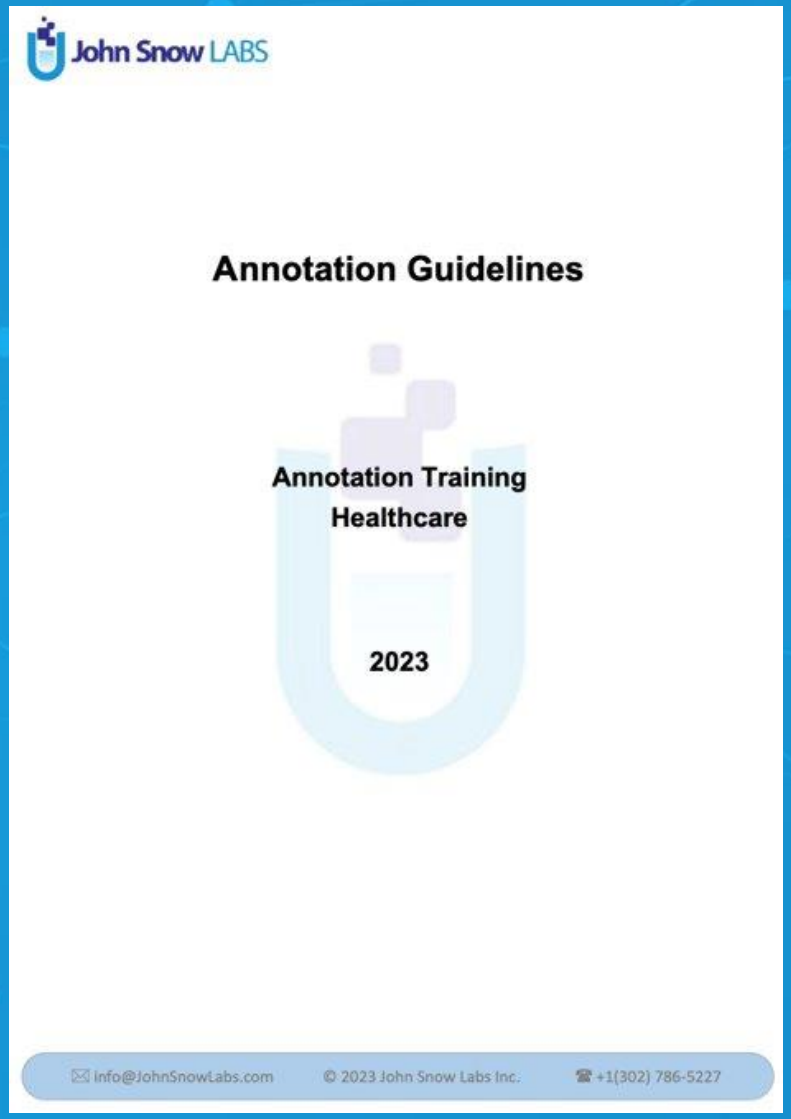
- ensure the quality and consistency of the annotations and AGs
- usually, an experienced annotator from the team is assigned as reviewer

John Snow LABS

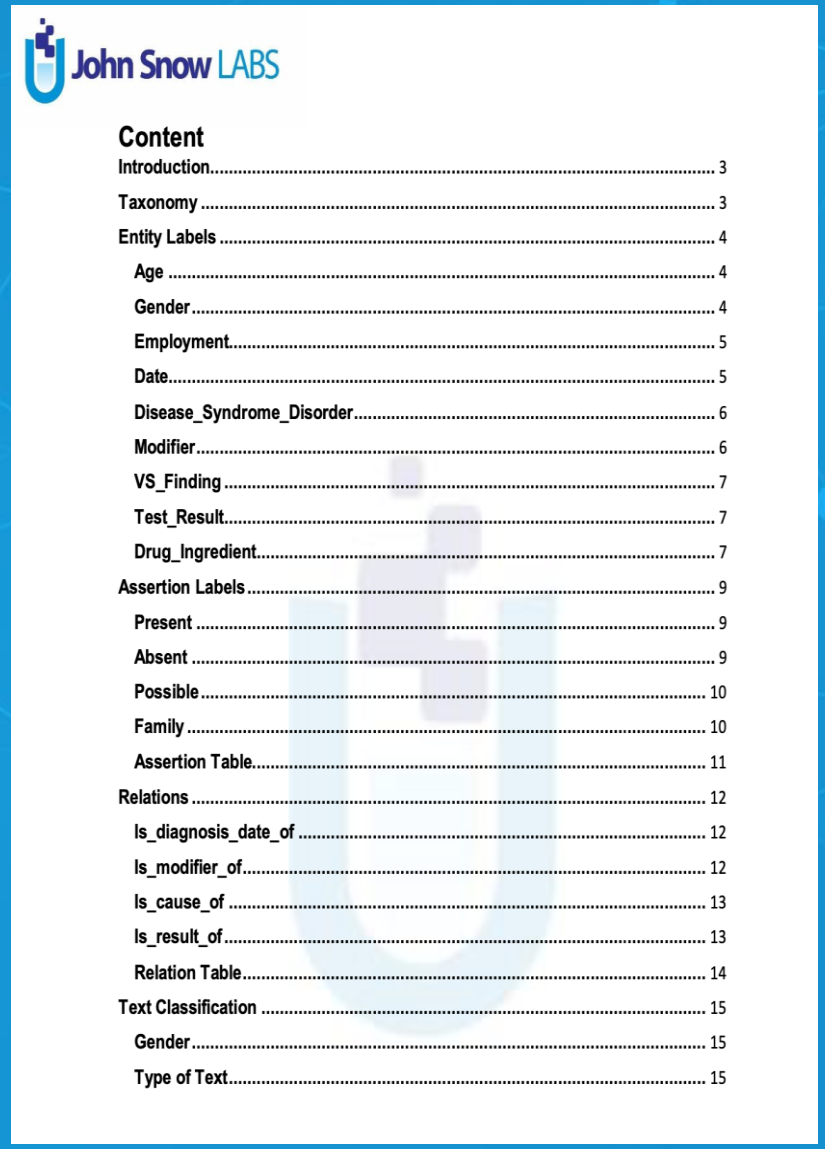
Structure and Content of AG

Contents

Contents



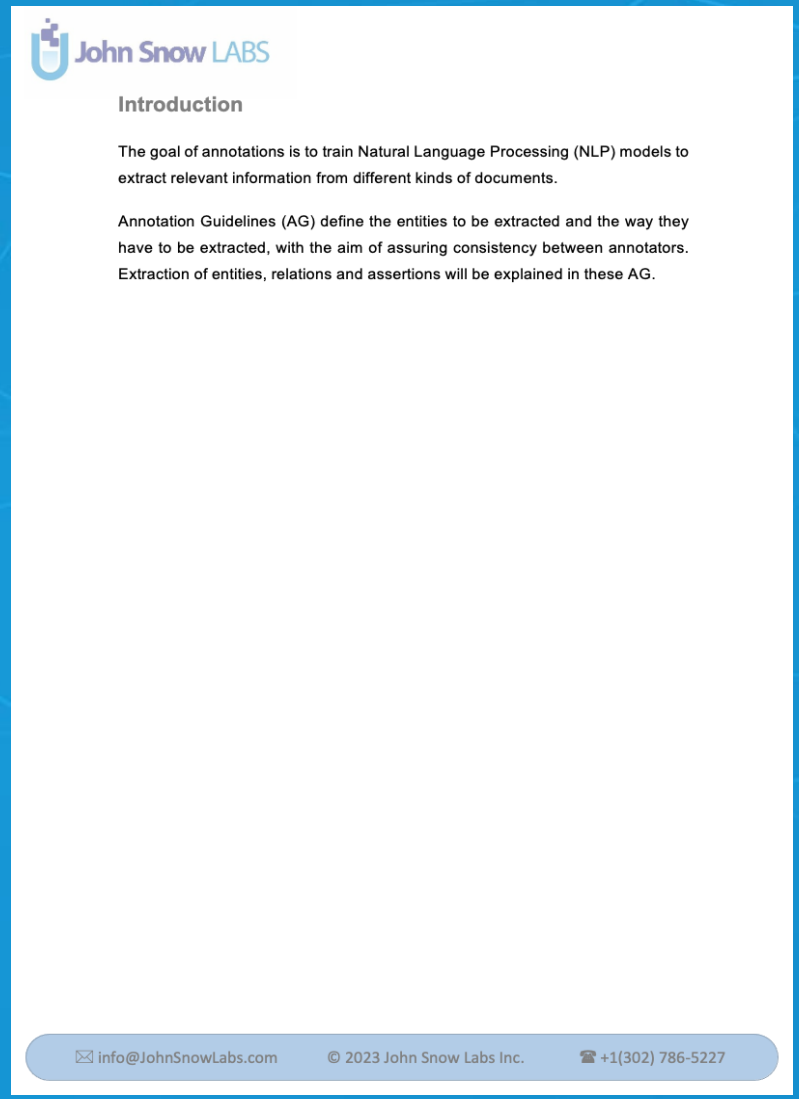
Contents



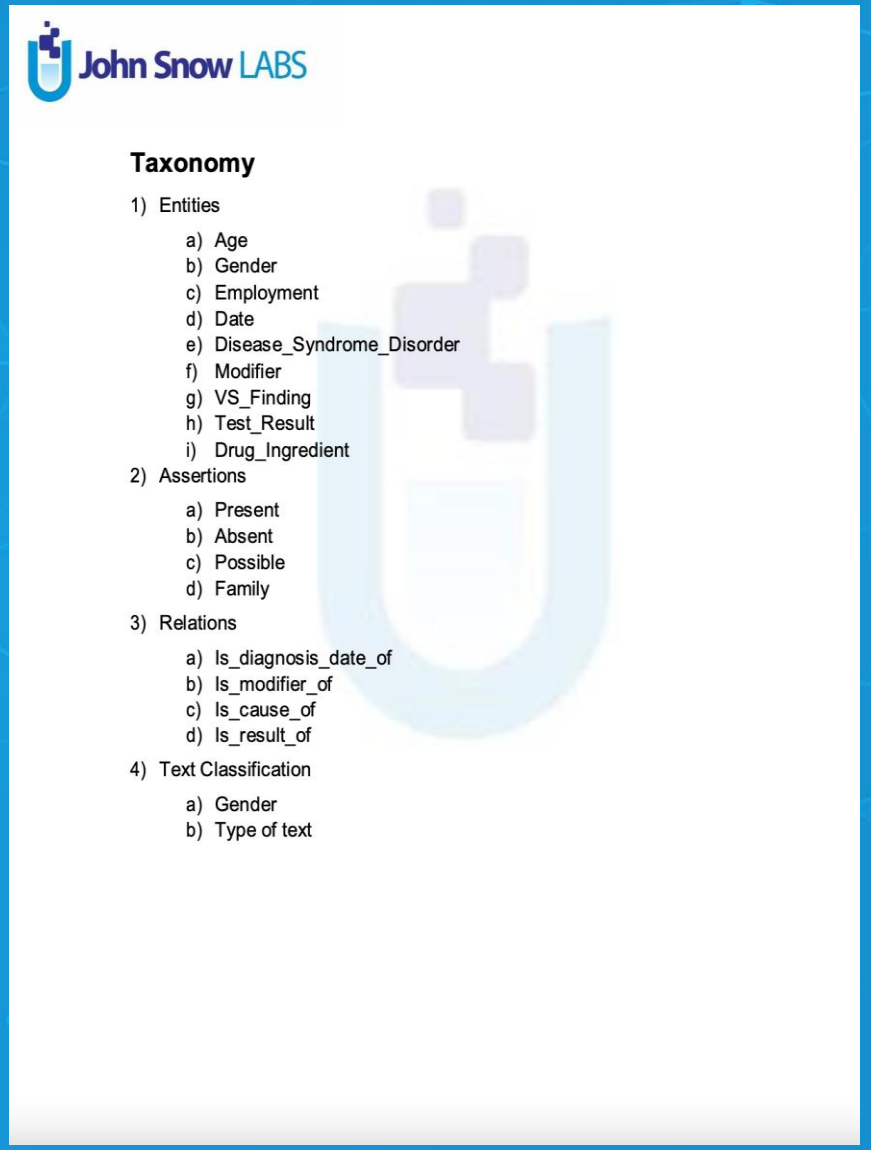
John Snow LABS

Content

Contents



Contents



John Snow LABS

Taxonomy

1) Entities
    a) Age
    b) Gender
    c) Employment
    d) Date
    e) Disease_Syndrome_Disorder
    f) Modifier
    g) VS_Finding
    h) Test_Result
    i) Drug_Ingredient
2) Assertions
    a) Present
    b) Absent
    c) Possible
    d) Family
3) Relations
    a) Is_diagnosis_date_of
    b) Is_modifier_of
    c) Is_cause_of
    d) Is_result_of
4) Text Classification
    a) Gender
    b) Type of text

Contents

John Snow LABS

Entity Labels

For each kind of entity, these AG provide a definition, specific extraction rules, examples, relations with other variables and assertion labels. Examples only include the annotations that are relevant for the kind of entity that is being explained. The following color reference is used:

- Green: Correct extraction.
- Grey: Incorrect extraction.

Every mention of a relevant entity must be extracted, even if it is repeated or negated. Punctuation marks should not be extracted unless stated otherwise. Extractions should neither start nor finish with a blank space.

Age

Definition: All mention of ages, past or present, related to the patient or with anybody else.
Extraction rules: Numbers should be extracted along with expressions such as "years old" or "age of". Ages expressed in terms of decades should also be extracted.
Examples:

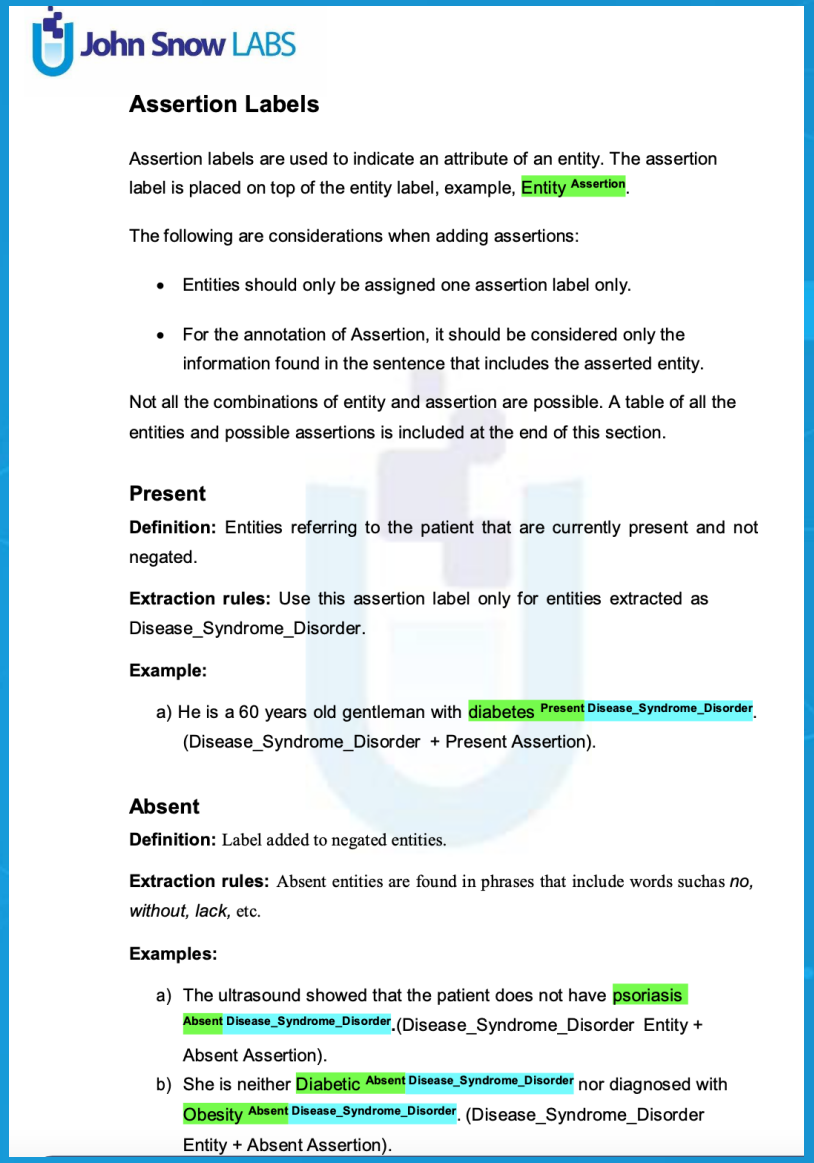
a) The patient presents as a 66-year-old Caucasian female in stable health.
b) The patient was diagnosed in his 50s.

Assertion labels : None.
Relations : None.

Gender

Definition: Gender-specific nouns, excluding family members.

Contents



John Snow LABS

Assertion Labels

Assertion labels are used to indicate an attribute of an entity. The assertion label is placed on top of the entity label, example, Entity Assertion.

The following are considerations when adding assertions:

- Entities should only be assigned one assertion label only.
- For the annotation of Assertion, it should be considered only the information found in the sentence that includes the asserted entity.

Not all the combinations of entity and assertion are possible. A table of all the entities and possible assertions is included at the end of this section.

Present

Definition: Entities referring to the patient that are currently present and not negated.

Extraction rules: Use this assertion label only for entities extracted as Disease_Syndrome_Disorder.

Example:

a) He is a 60 years old gentleman with diabetes Present Disease_Syndrome_Disorder. (Disease_Syndrome_Disorder + Present Assertion).

Absent

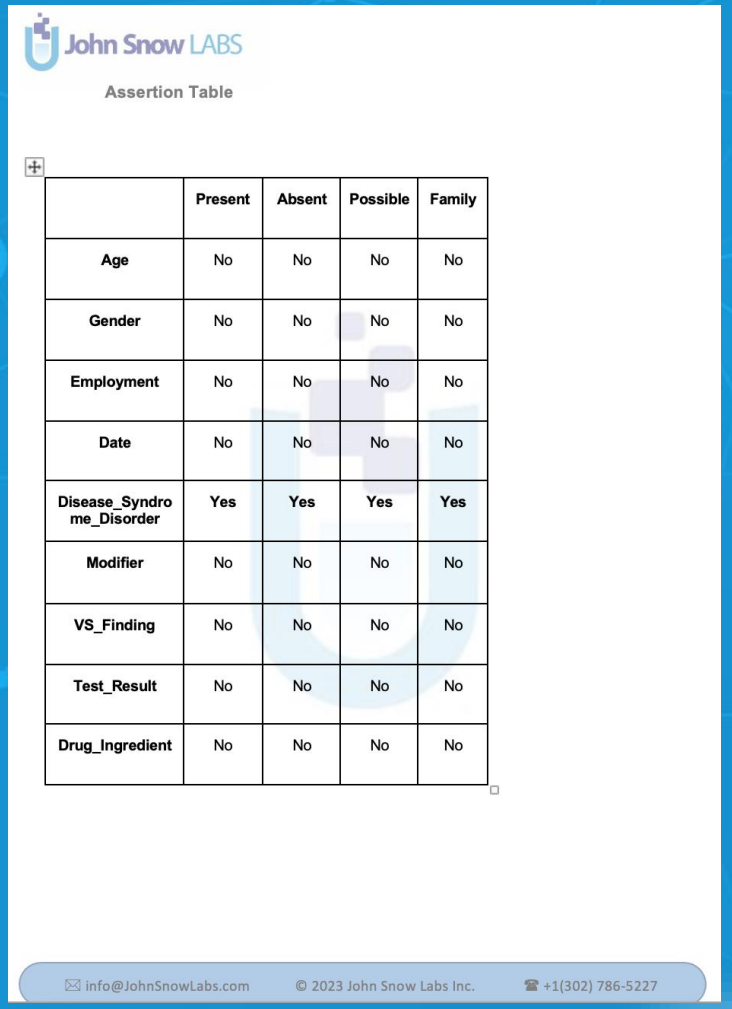
Definition: Label added to negated entities.

Extraction rules: Absent entities are found in phrases that include words suchas *no, without, lack,* etc.

Examples:

a) The ultrasound showed that the patient does not have psoriasis Absent Disease_Syndrome_Disorder.(Disease_Syndrome_Disorder Entity + Absent Assertion).

b) She is neither Diabetic Absent Disease_Syndrome_Disorder nor diagnosed with Obesity Absent Disease_Syndrome_Disorder. (Disease_Syndrome_Disorder Entity + Absent Assertion).



John Snow LABS

Assertion Table

	Present	Absent	Possible	Family
Age	No	No	No	No
Gender	No	No	No	No
Employment	No	No	No	No
Date	No	No	No	No
Disease_Syndrome_Disorder	Yes	Yes	Yes	Yes
Modifier	No	No	No	No
VS_Finding	No	No	No	No
Test_Result	No	No	No	No
Drug_Ingredient	No	No	No	No

info@JohnSnowLabs.com © 2023 John Snow Labs Inc. +1(302) 786-5227

Contents

John Snow LABS

Relations

Relations are used to link two related entities. To create relations between entities, use the Create Relation button of the annotation tool. Relations are **NOT** created for entities present in different sentences, or are 2 or more sentences apart. Some relations require a relation label that is found in the relation section of the annotation tool. Also, some relations require assignation of direction that is represented by an arrow in the relation section of the annotator tool. A table with all the possible relations is included at the end of this section.

Is_diagnosis_date_of

Definition: This relation is used to associate a Disease_Syndrome_Disorder entity and a Date entity.

Extraction rules: The Disease_Syndrome_Disorder entity and the relevant date associatedwith it are extracted and related using the relation label is_diagnosis_date_of only when the date refers to the diagnosis of the medical problem.

Examples:

a) She was diagnosed with hypertension in 1987. **Hypertension** | (Disease_Syndrome_Disorder entity) and **1978** (Date entity) are related with is_diagnosis_date_of label.

Is_modifier_of

Definition: This relation is used to associate a Disease_Syndrome_Disorder entity and aModifier.

Extraction rules: The Disease_Syndrome_Disorderand the relevant modifier are extractedand related using the relation label is_modifier_of.

Examples:

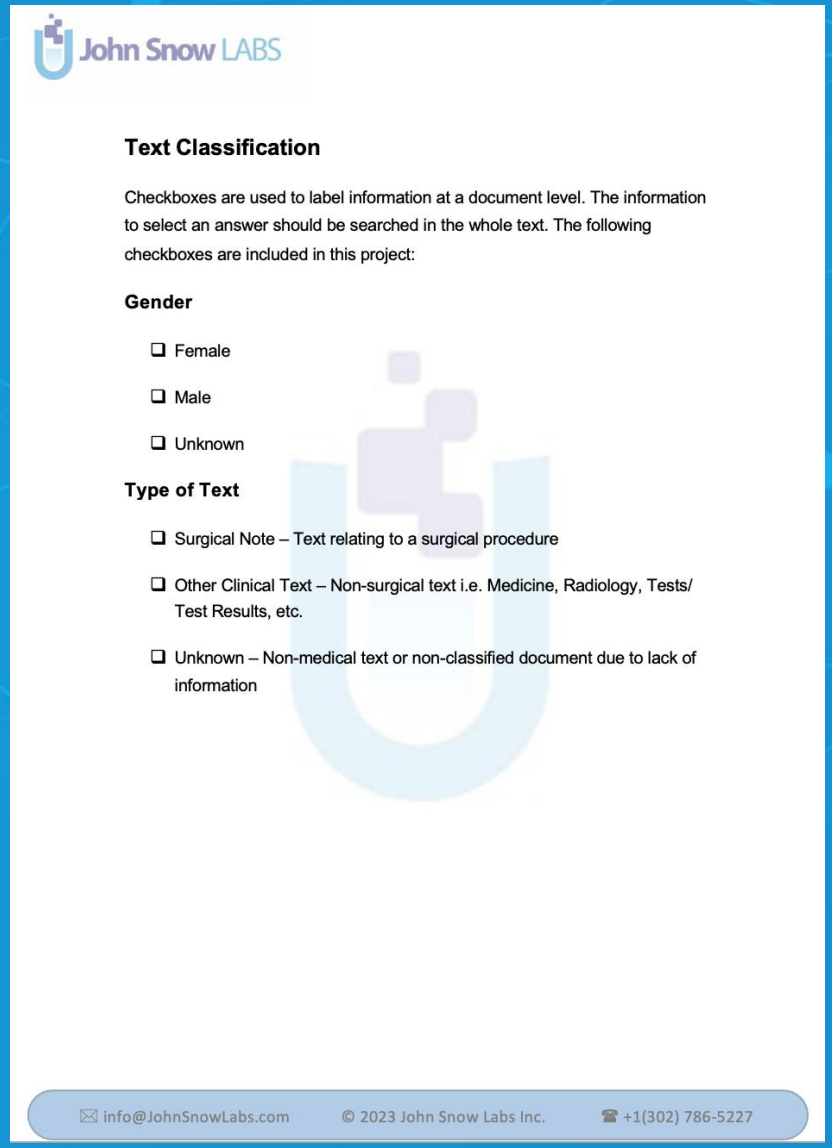
a) He has been experiencing chronic migraine for five years. **Migraine** (Disease_Syndrome_Disorderentity) and **Chronic** (Modifier entity) are related with is_modifier_of label.

John Snow LABS

Relation Table

	Entity 1	Entity 2	Label Needed	Direction Needed
Is_diagnosis_date_of	Disease_Syndrome_Disorder	Date	Yes	No
Is_modifier_of	Disease_Syndrome_Disorder	Modifier	Yes	No
Is_cause_of	Disease_Syndrome_Disorder	Disease_Syndrome_Disorder	Yes	Yes
Is_result_of	VS_Finding	Test_Result	Yes	No

Contents



John Snow LABS

Text Classification

Checkboxes are used to label information at a document level. The information to select an answer should be searched in the whole text. The following checkboxes are included in this project:

Gender

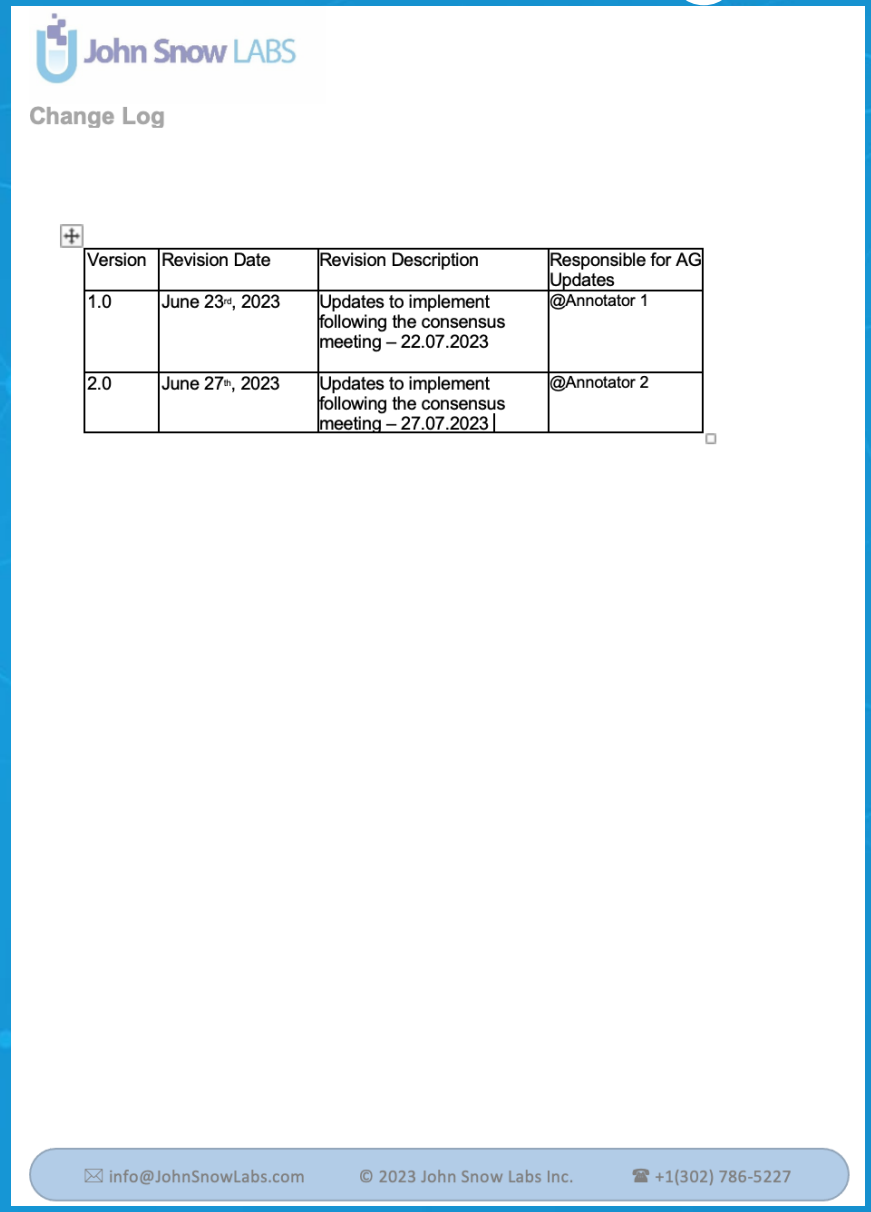
- ❑ Female
- ❑ Male
- ❑ Unknown

Type of Text

- ❑ Surgical Note – Text relating to a surgical procedure
- ❑ Other Clinical Text – Non-surgical text i.e. Medicine, Radiology, Tests/ Test Results, etc.
- ❑ Unknown – Non-medical text or non-classified document due to lack of information

✉ info@JohnSnowLabs.com © 2023 John Snow Labs Inc. ☎ +1(302) 786-5227

Contents



John Snow LABS

Change Log

Version	Revision Date	Revision Description	Responsible for AG Updates
1.0	June 23rd, 2023	Updates to implement following the consensus meeting – 22.07.2023	@Annotator 1
2.0	June 27th, 2023	Updates to implement following the consensus meeting – 27.07.2023	@Annotator 2

Examples #1

Age

Definition: All mention of ages, past or present, related to the patient or with anybody else.

Extraction rules: Numbers should be extracted along with expressions such as "years old" or "age of". Ages expressed in terms of decades should also be extracted.

Examples:

a) The patient presents as a 66-year-old Caucasian female in stable health.

b) The patient was diagnosed in his 50s.

c) John is a 65 years old male.

d) He was diagnosed with obesity at the age of 25.

Assertion labels : None.

Relations : None.

Examples #2

Employment

Definition: Mentions of jobs or occupations included in the text.

Extraction rules: Extract terms that are related to any specific jobs or employment,whether related to the patient or not. Do not extract words such as "works", "working" or "employed".

Examples:

a) She is an office manager for a gravel company.
b) She will also see a nutritionist and a social worker.
c) He works as a financial officer.

Assertions: None.

Relations: None.

Examples #3

Disease_Syndrome_Disorder

Definition: Extract all the diseases, syndromes and any relevant condition mentioned in the document.

Extraction rules: Extract all mentions of medical conditions and diseases, including those related to the patient or to a family member. Do not include in the extraction modifiers such as "chronic", "mild" or "severe" (this kind of words should be extracted using the label Modifier).

Examples:

a) The patient has Alzheimer diagnosed back in 2012.
b) He was diagnosed with colon cancer.
c) A diagnosis of chronic kidney disease was established in the past.
d) He was diagnosed twice with chronic depression.

Modifier

Definition: Terms that modify the medical problem.

Extraction rules: Extract words that indicate severity (such as "mild" or "severe"), duration (such as "chronic" or "acute") or any other feature of the entities.

Examples:

a) He has been experiencing chronic back pain for five years.
b) Patient with history of recurrent angina.

John Snow LABS

Best Practices

Best Practices - Annotation Rules

Centralized

one document should include all the rules

(the AGs)

Consistent

contradiction should be avoided

Specific

ambiguity should be avoided

Cross-check Annotations

Use multiple annotators for the same data

(“seed corpus”)

Clear Objectives

objectives will shape your annotation guidelines

Explicit

i.e written (even if in detail!)

Best Practices - AG Development

Define goals clearly

What information is needed from texts?

e.g.: “identifying all gender-related words (NER)”

vs.

“classifying the document based on gender” (Text Classification +/- NER)

Communication is key

Define communication channels (Slack, Teams etc.)

Keep track of decisions made (Change Log)

Encourage annotators to avoid private messaging

Iterative process

Test your AGs

Seed Corpus annotations (pilot)

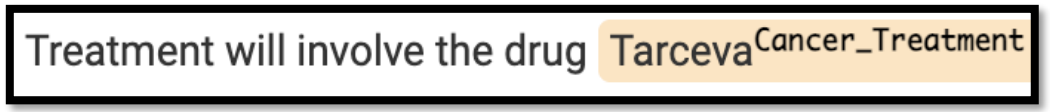
data scientists/NLP engineers

Best Practices - NER

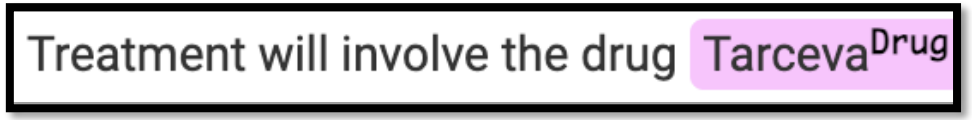
vs

Define the level of **granularity** of your taxonomy

e.g.: ***Clinical_Problem*** label vs. ***Modifier + Symptom***



vs



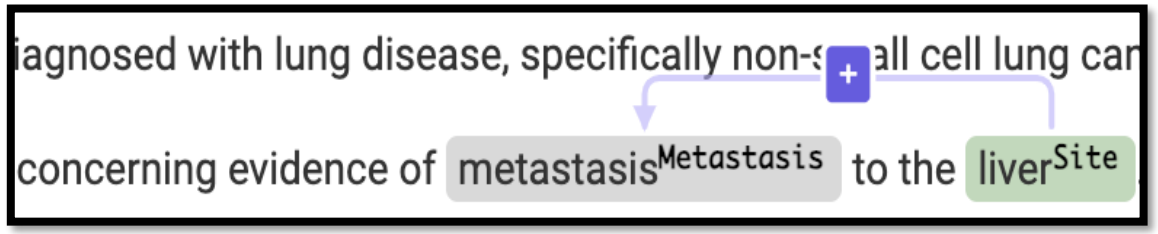
Each label should have **clear boundaries** (avoid overlapping)

e.g.: “Treatment will involve the drug Tarceva”

(***Cancer_Treatment*** vs ***Drug***)

Best Practices - NER

vs



Extractions should not include more than 2-3 words!

Consider merging/dropping/splitting entities

e.g.: "There is metastasis in the lung" (Metastasis > Metastasis + Site)

Combine NER with **other NLP features**

assertion labels

relations

text classification

Best Practices - Assertion

Keep **Assertion Taxonomy** simple!

vs

- Assign a default assertion status

e.g.: "The patient has cancer". (Disease + Present assertion label by Default)

- Define each assertion label

How to decide between multiple possible assertion labels

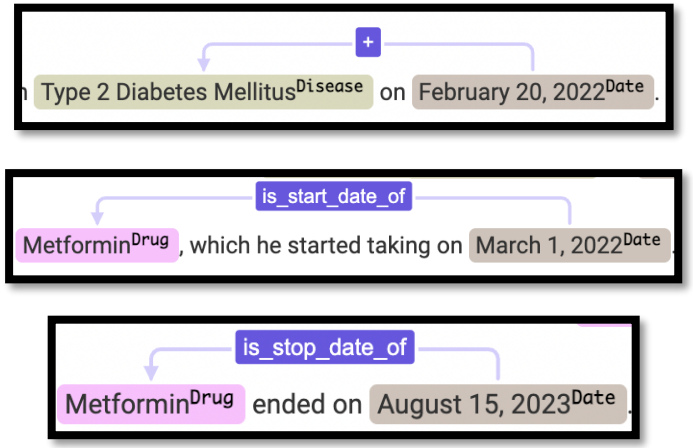
e.g.: "No family history of cancer". (Disease + Absent/Family?)

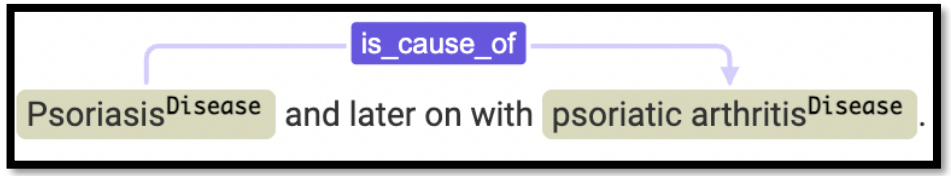
Create your own assertion labels

e.g.: "medical history reveals an allergy to penicillin" (Drug NER label + Allergen assertion label)

Best Practices - Relations

Define **entities** that can be related!





- Add Relation Labels Names only when there are different ways of relating two entities

Disease & Date (1 relation): no need to add a Label is_date_of_diagnosis

- ***Drug & Date (3 relations): label needed (is_start_date_of, is_stop_date_of, is_generic_date_of)***

- Add direction only if necessary

is_cause_of (Disease & Disease) needs direction to differentiate cause and consequence

- Define the **distance between entities** to use relation label

John Snow LABS

Frequent Annotation Errors

Cause of Error

1. Accidental Human Error

Example: “female" extracted as Age

2. Wrong Extraction

Example: “young" extracted as Age (if AG specify that only numeric values should be extracted).

3. Normal Disagreement

Example: different granularity to extract “late 40s“ as Age.

Kinds of NER Errors

1. Missing Extraction

"The patient is 30 years old" [O]

2. Wrong Label

"The patient is 30 years old" [Gender]

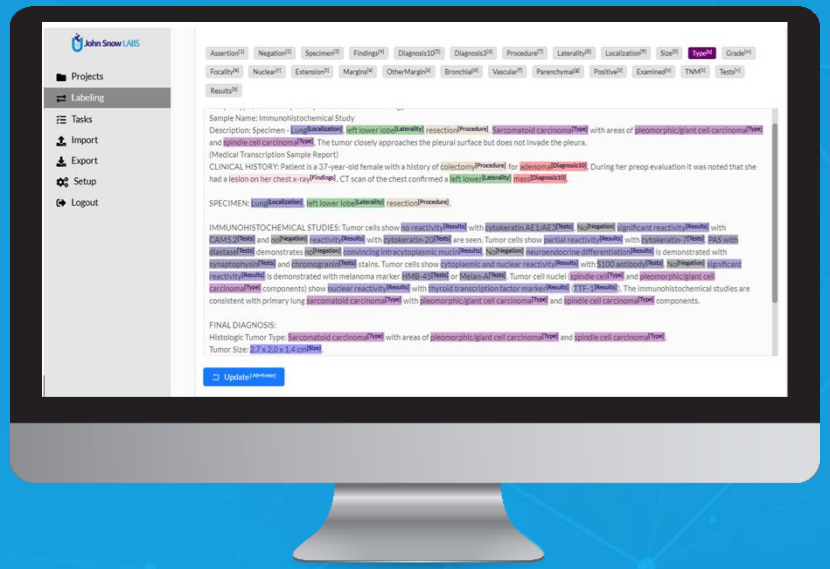
3. Wrong Granularity

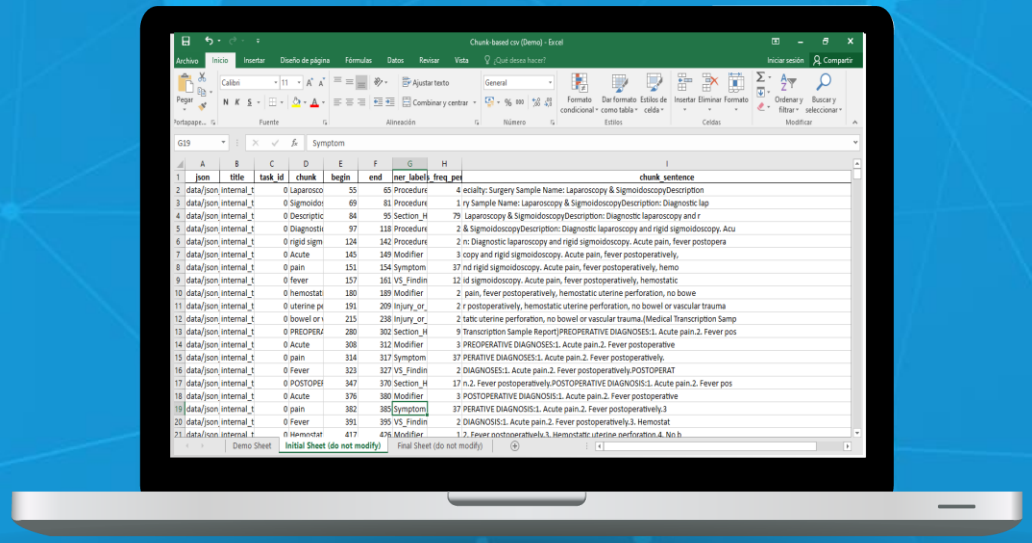
"The patient is 30 years old" [Age]

4. Wrong Extraction

"The patient is young" [Age] (redundant if AG specify "only numeric values")

Data Review





Manual Reviews

CSV Reviews

John Snow LABS

Project Analytics – good example

Exercise – NER Annotation following AG

1. Create new project with the configuration specified here
2. Import tasks from json file
3. Assign all tasks to you as annotator
4. Annotate tasks according to AGs
5. Submit completion as Ground Truth (starred)

Outline

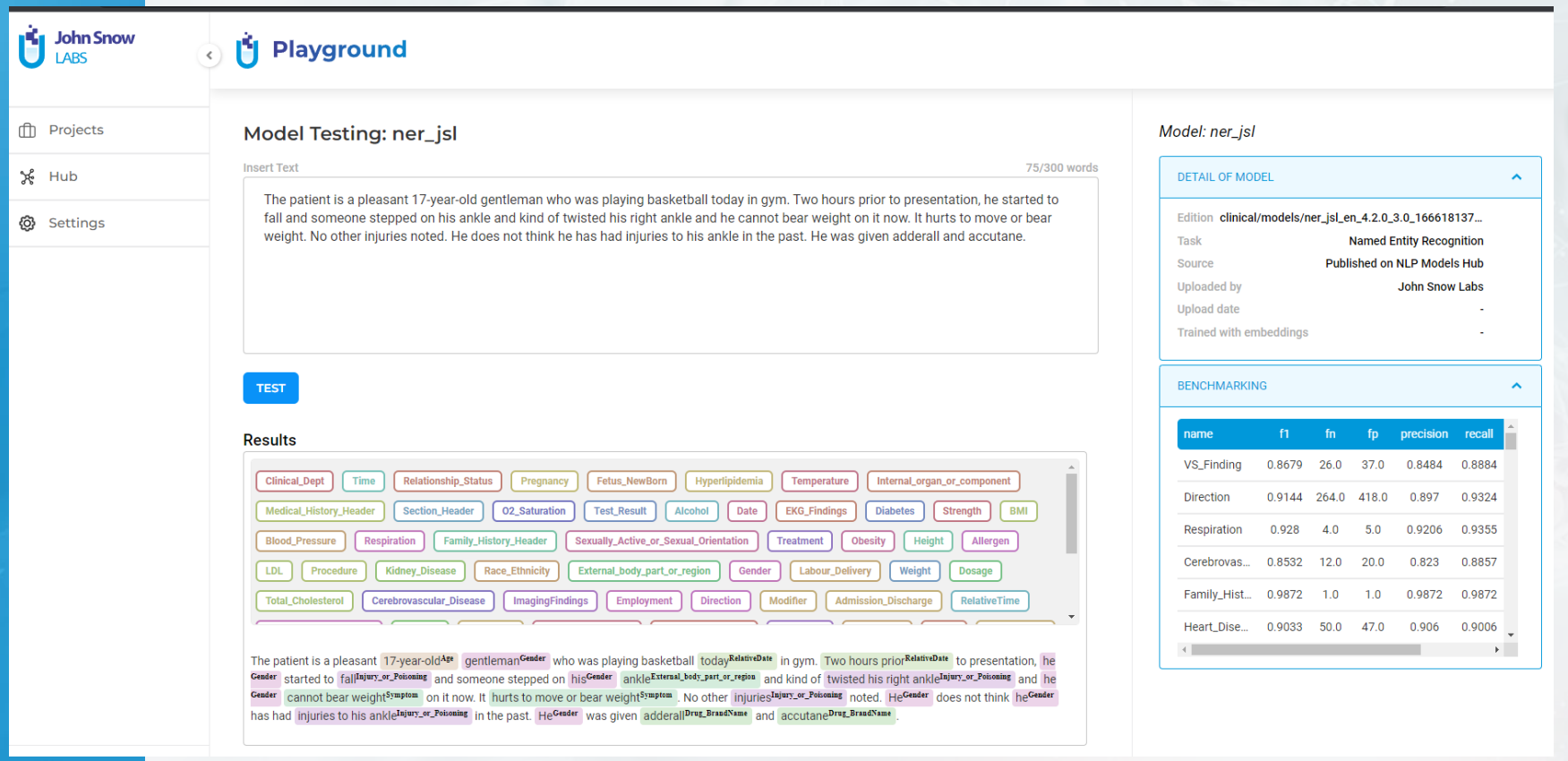
1. Introduction to Text Annotation
2. Annotation Projects Setup in NLP Lab
3. Annotation Guidelines
4. **Preannotation Resources**
5. Model Training
6. Conclusions and further resources

The Hub Of Resources

1. **Integration with NLP Models Hub**
   - ✓ Check benchmarking information
   - ✓ Download any resource with one click
2. **Models**
   - ✓ Private repository of models
   - ✓ Trained with NLP Lab, Manually uploaded, Downloaded from NLP Models Hub
3. **Rules (examples available here)**
   - ✓ Private repository of rules
   - ✓ Rules editing interface
4. **Prompts (examples available here)**
   - ✓ Private repository of prompts
   - ✓ Prompt editing interface

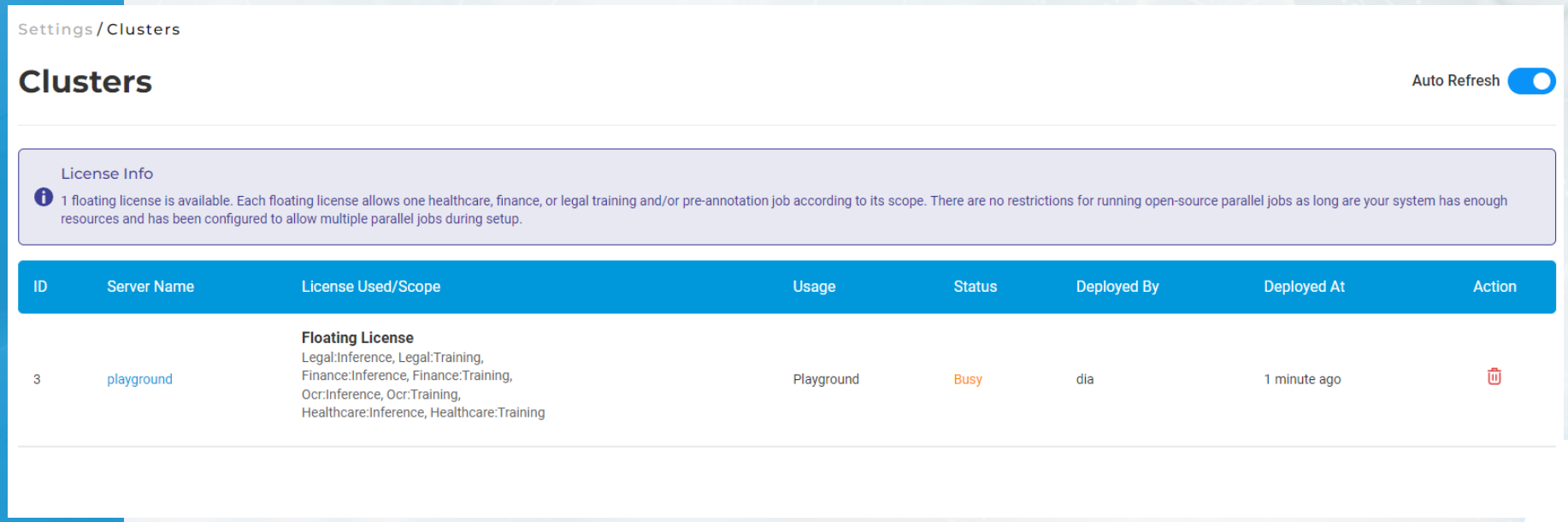
Playground

1. Deploy resource for quick tests on custom documents
2. Edit rules and prompts on the fly

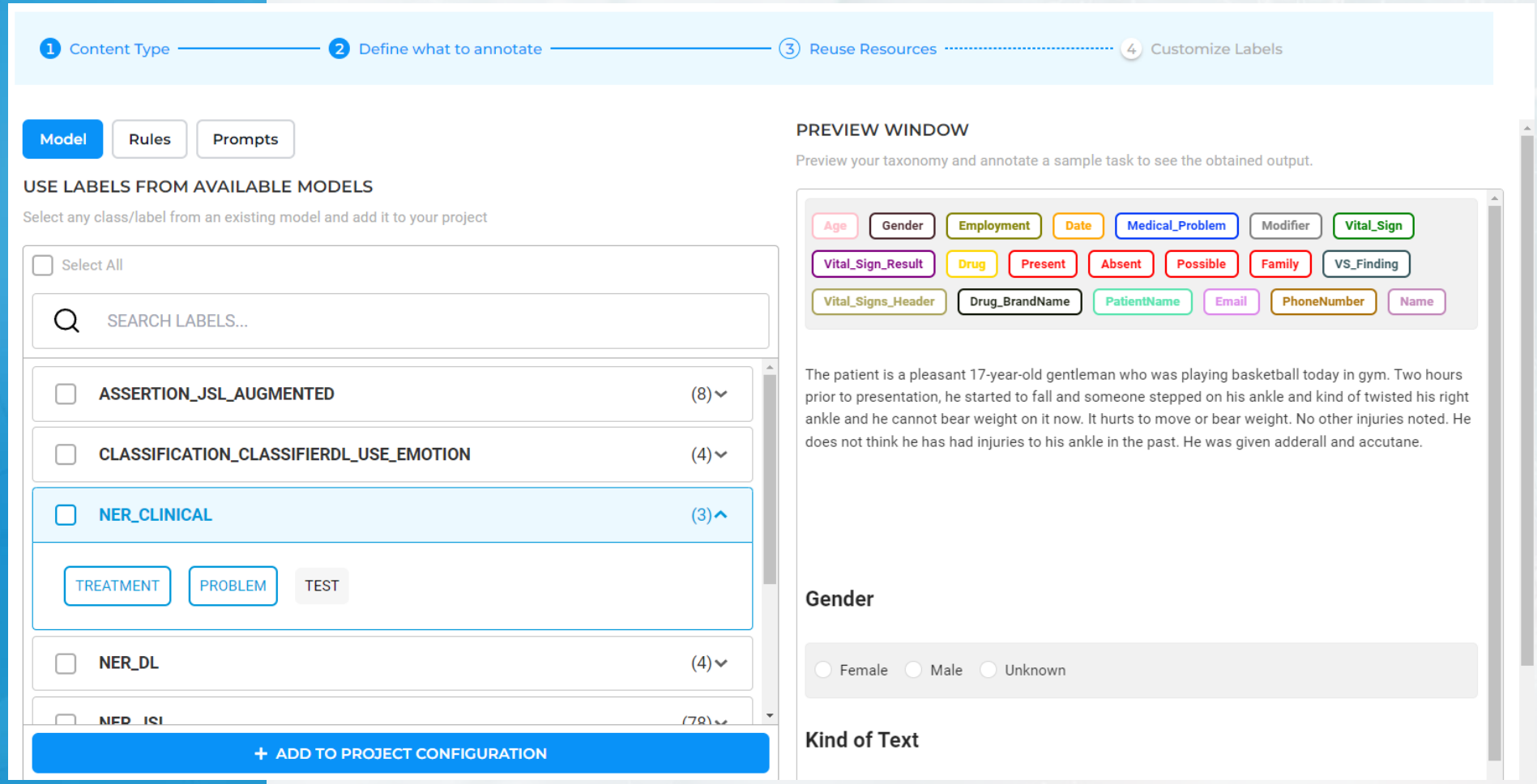


Clusters

1. Up-to-date information on the deployed servers
2. Computation resource management
3. Access deployed playground



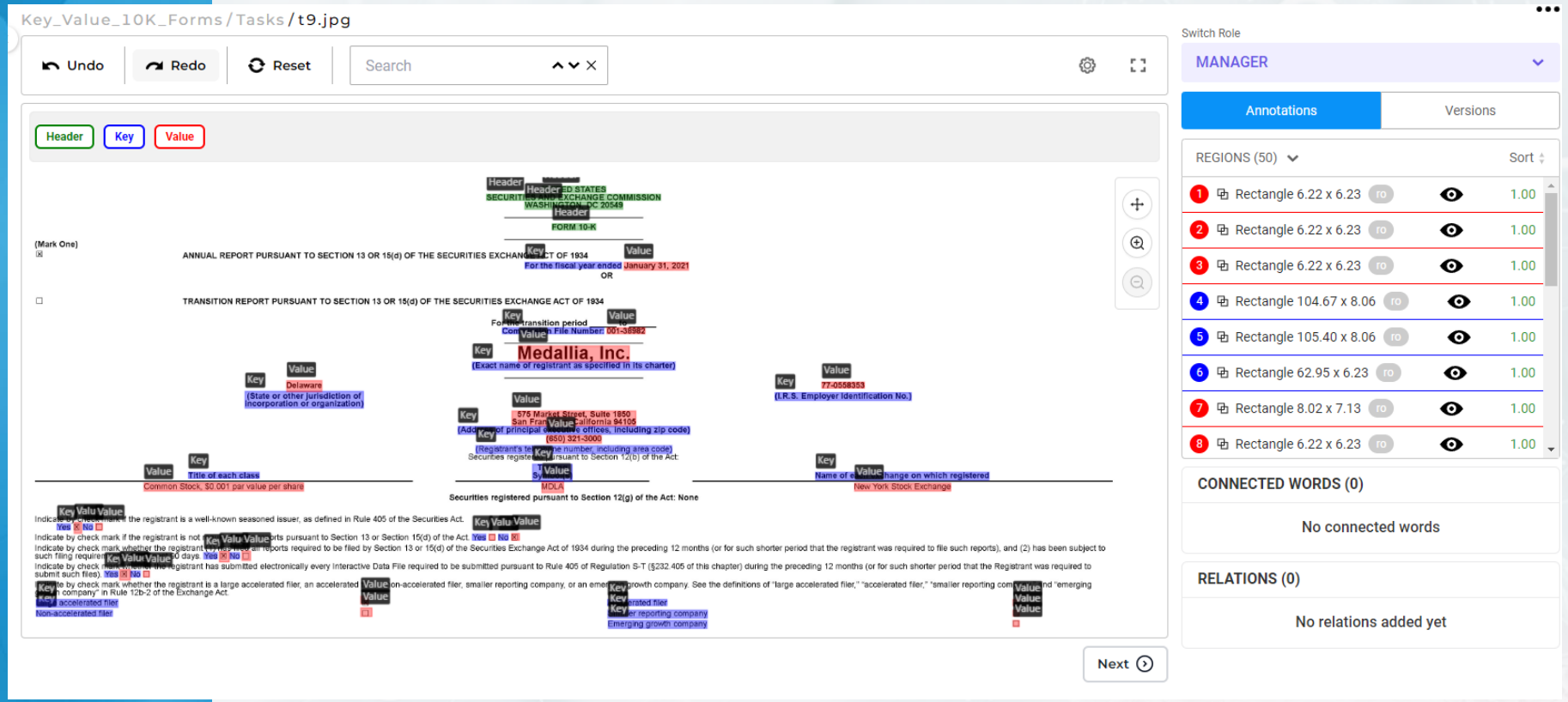
Reuse Resources in your projects



Exercise – Clinical Data Annotation

1. Create a project with the configuration described here.
2. Assign all tasks to you as annotator.
3. Run preannotations on all tasks.
4. Click on one task to access the annotation screen.
5. Check the pre-annotations + click Edit and if necessary, update the annotations according to your AG.
6. Save your work and submit the new version of annotations.

Support for PDF and Image



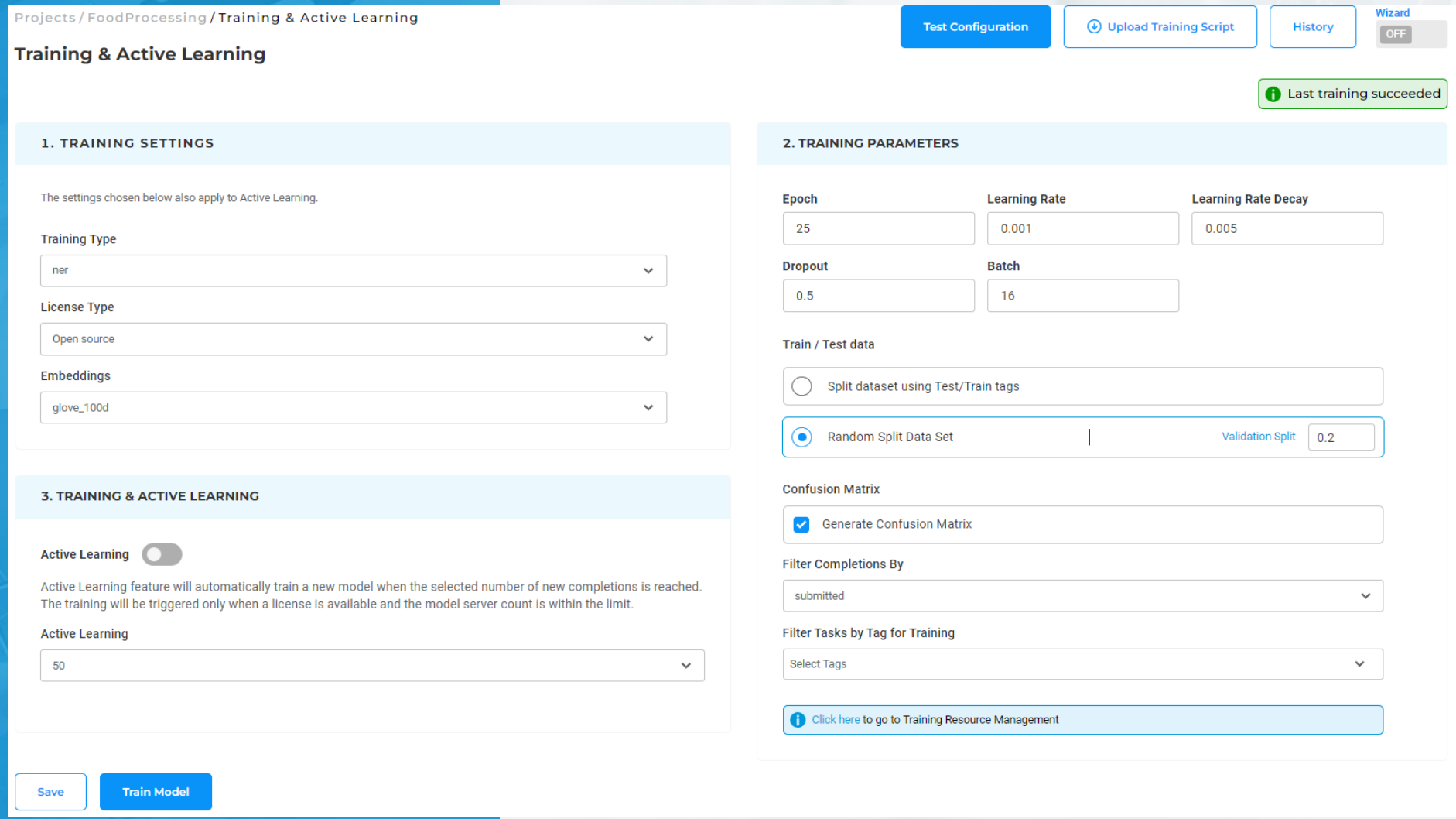
- OCR license required
- Select Visual NER template
- Create OCR server for pdf/image import

- Sample configuration
- Sample tasks

Outline

1. Introduction to Text Annotation
2. Annotation Projects Setup in NLP Lab
3. Annotation Guidelines
4. Preannotation Resources
5. **Model Training**
6. Conclusions and further resources

Model Training



3 easy steps:

1. Choose the model type and the embeddings
2. Set training params including the tasks to use
3. Optional – turn on the active learning for future trainings

Example project

- Food Processing

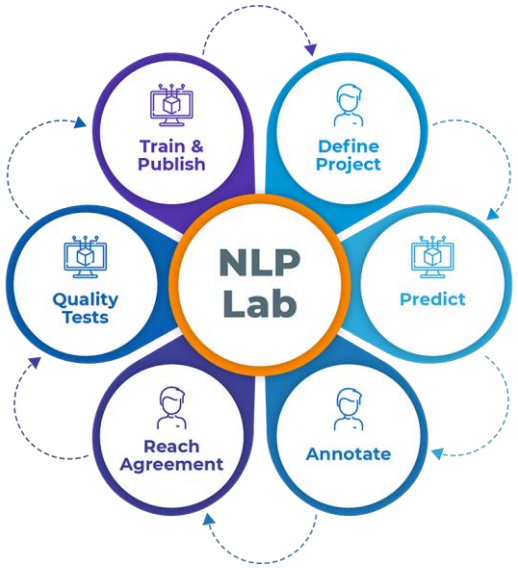
Outline

1. Introduction to Text Annotation
2. Manual Annotation
3. Annotation Guidelines
4. Project Setup and Management
5. Preannotations
6. **Conclusions and further resources**

Conclusions – NLP Lab Facilitates

1. Manual annotation process
2. Best practices for efficient and coherent annotations
3. Project templates and project configuration to suite your project
4. Reuse pretrained resources for faster and more efficient results
5. Detect annotation errors with analytics
6. Train models to learn common tasks

Learning Resources



Docs https://nlp.johnsnowlabs.com/docs/en/alab/quickstart

Tutorials https://nlp.johnsnowlabs.com/docs/en/alab/step_by_step_tutorials

Blog https://www.johnsnowlabs.com/nlp-lab-blog/

Quick Install

- AWS Marketplace
- Azure Marketplace
- OCI Marketplace
- On-Premise

Support

- Slack
- Email

Questions?

- dia@johnsnowlabs.com
- andrei@johnsnowlabs.com
- radu@johnsnowlabs.com