

Hands-On-Session 1: Medical Report Data Analysis – practical

Timetable, Label Studio, Google Colab

Hands-on Session Organization

The Rules

- Rule 0:
 - You are not alone. I'm here to help.
 - **Don't panic!**
- Rule 1:
 - Everything (slides, data, links, explanations) can be found here:
 - <https://github.com/iml-r/heidelberg-nmt>



Timetable

- Theoretical presentation (the one before; 10 minutes)
- Practical presentation (the one now; 10 minutes)
- Annotation block (30-50 minutes)
 - Getting acquainted with Label Studio
 - Our browser-based annotation environment
 - Annotating
 - Discussion
- Programming/analysis block (20-25 minutes)
 - Coding in Colab – annotation evaluation
 - Discussion and specific cases deep dive

Total:

90 minutes tops

Label Studio

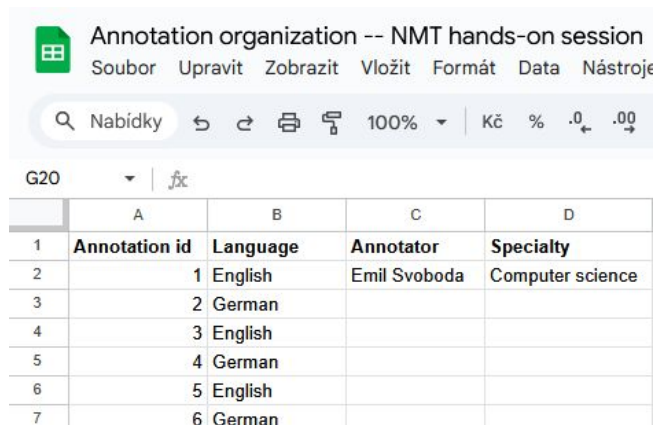
Label Studio

- An open-source annotation environment
- We run it on our hardware – third parties (Google, OpenAI)
- You can join it just by clicking this link (you'll have to register):
 - <https://quest.ms.mff.cuni.cz/workshop-heidelberg/>
- After you've registered and logged in, you should see something like this:



Organization sheet – BEFORE YOU START ANNOTATING

- **!!! WE NEED TO ENSURE THAT EACH TASK IS ANNOTATED BY ONE PERSON AND ONE PERSON EXACTLY !!!**
- So pick a task from the (publicly editable) Google spreadsheet:
 - <https://docs.google.com/spreadsheets/d/17H3dxLnWugpMYFqhGrSz3bBfefHUCPy916EZSaovrA/edit?usp=sharing>
- Pick a task by id and language, fill out your name (you can make up one up, but use the same one for each task) and specialty (computer science vs. medicine)



The screenshot shows a Google Spreadsheet interface. The title bar reads 'Annotation organization -- NMT hands-on session'. Below the title bar is a menu bar with options: Soubor, Upravit, Zobrazit, Vložit, Formát, Data, and Nástroje. Below the menu bar is a search bar with the text 'Nabídky' and a magnifying glass icon. To the right of the search bar are icons for undo, redo, print, and a zoom dropdown set to 100%. Further right are currency symbols: Kč, %, and a decimal separator icon. Below the search bar is a table with 4 columns: A, B, C, and D. The table has 7 rows. The first row is a header row with the following text: '1', 'Annotation id', 'Language', 'Annotator', and 'Specialty'. The second row contains: '2', '1', 'English', 'Emil Svoboda', and 'Computer science'. The third row contains: '3', '2', 'German', and empty cells for 'Annotator' and 'Specialty'. The fourth row contains: '4', '3', 'English', and empty cells for 'Annotator' and 'Specialty'. The fifth row contains: '5', '4', 'German', and empty cells for 'Annotator' and 'Specialty'. The sixth row contains: '6', '5', 'English', and empty cells for 'Annotator' and 'Specialty'. The seventh row contains: '7', '6', 'German', and empty cells for 'Annotator' and 'Specialty'.

	A	B	C	D
1	Annotation id	Language	Annotator	Specialty
2		1 English	Emil Svoboda	Computer science
3		2 German		
4		3 English		
5		4 German		
6		5 English		
7		6 German		

Label Studio

Welcome 🙌

Let's get you started.

Create Project

Invite People

Recent Projects

Hands-on German-to-English NMT evaluation
0 of 2130 Tasks (0%)

[View All](#)

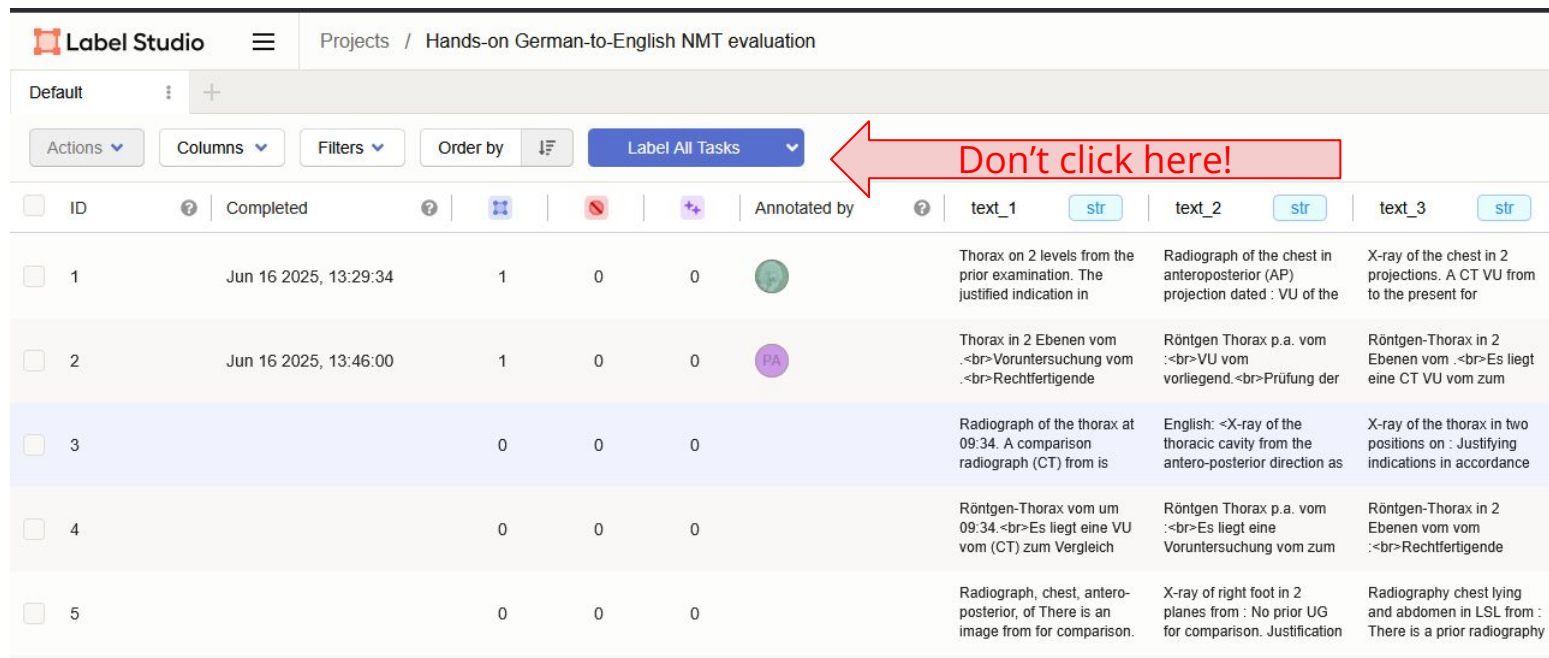
Do click here!

DON'T CLICK
HERE*

*if you do, just hit F5

Label Studio

Always reserve a task first in the organization sheet – the “Annotated by” only appears after the task has been completed – and DON’T click “Label All Tasks”



The screenshot shows the Label Studio interface for a project named "Hands-on German-to-English NMT evaluation". The interface includes a top navigation bar with the Label Studio logo and a sidebar with a menu icon. The main area displays a table of tasks. A red arrow points to the "Label All Tasks" button, which is highlighted in blue. The table has columns for ID, Completed, and Annotated by, among others. The tasks are listed with their respective IDs, completion times, and annotations.

ID	Completed	Annotated by	text_1	text_2	text_3
1	Jun 16 2025, 13:29:34		Thorax on 2 levels from the prior examination. The justified indication in	Radiograph of the chest in anteroposterior (AP) projection dated : VU of the	X-ray of the chest in 2 projections. A CT VU from to the present for
2	Jun 16 2025, 13:46:00		Thorax in 2 Ebenen vom : Voruntersuchung vom : Rechtfertigende	Röntgen Thorax p.a. vom : VU vom vorliegend. Prüfung der	Röntgen-Thorax in 2 Ebenen vom : Es liegt eine CT VU vom zum
3			Radiograph of the thorax at 09:34. A comparison radiograph (CT) from is	English: <X-ray of the thoracic cavity from the antero-posterior direction as	X-ray of the thorax in two positions on : Justifying indications in accordance
4			Röntgen-Thorax vom um 09:34. Es liegt eine VU vom (CT) zum Vergleich	Röntgen Thorax p.a. vom : Es liegt eine Voruntersuchung vom zum	Röntgen-Thorax in 2 Ebenen vom vom : Rechtfertigende
5			Radiograph, chest, antero-posterior, of There is an image from for comparison.	X-ray of right foot in 2 planes from : No prior UG for comparison. Justification	Radiography chest lying and abdomen in LSL from : There is a prior radiography

- Then, you'll be presented with a sequence of 8 texts.
- Select 1 (or more) diagnoses for each text.
- Tick only the diagnoses that are *new findings*; e.g. don't annotate *Wrist fracture* if the report is from a follow-up exam.

CXR PA of 15:17:
Existing CXR from the same day at 10:07 is provided as a comparison.
Right Pleur-X drainage recently installed for the CXR with the tip in the projection of the right basal area, with slight regression of homogeneous opacity reduction on the right with residual findings. No clear pleural dehiscence. Unchanged heart/mediastinal contours. No new confluent infiltrate. Fringe-like condensation in the projection of the right basal area. Free air infradiaphragmal right with the formation of a gas-fluid level, from CT CXR of discovered lesion formation. Operative clip material in the projection of the liver shadow after right hemihepatectomy. Currently installed PTCd from the left. Soft tissues and bony structures are normal. ||| No indication of pneumothorax with the right pleura catheter, moderate residual effusion. Improved ventilation of the right basal area with striated ventilation disorders remaining. Apart from this, no significant change in findings compared to the previous follow-up. | No indication of pneumothorax with the right pleura catheter, moderate residual effusion. Improved ventilation of the right basal area with striated ventilation disorders remaining. Apart from this, no significant change in findings compared to the previous follow-up. ||| Pleural catheter inserted without complications on the right.

☐ Pneumothorax ☐ Fracture of unspecified body region ☐ Pneumonia

☐ Other specified functions of the cardiovascular, haematological, immunological and respiratory systems ☐ Fracture of forearm

☐ Fracture of shoulder or upper arm ☐ Fracture of rib, sternum or thoracic spine ☐ Fracture of femur

☐ Fracture of lower leg, including ankle ☐ Fracture of lumbar spine or pelvis ☐ Fracture at wrist or hand level

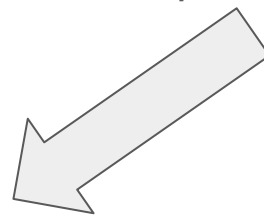
☐ Cardiomegaly ☐ Fluid overload ☐ Pleural effusion ☐ Low bone mass disorders ☐ Emphysema

☐ Traumatic subcutaneous emphysema, not elsewhere classified ☐ Other ☒ Healthy

⏮ ⏪ ✕ ⚙

Submit

Always click submit
after you've
annotated your 8
reports

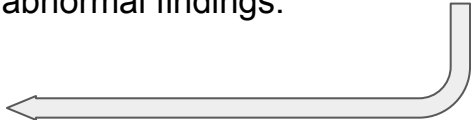


Label Studio – label set

Condition

Description

Pneumothorax	Air leaks into the space around the lungs, possibly collapsing part or all of a lung.
Fracture of unspecified body region	A bone break is detected, but the location is not listed in the other labels (e.g. facial fracture)
Cardiomegaly	Enlargement of the heart, often linked to heart disease or chronic high blood pressure.
Fluid overload	Excess fluid in the body, often affecting the lungs, due to kidney or heart failure.
Pleural effusion	Fluid buildup in the space between the lungs and chest wall.
Low bone mass disorders	Conditions like osteoporosis that weaken bones, increasing fracture risk.
Emphysema	Chronic lung damage where air sacs are destroyed, making it hard to breathe.
Traumatic subcutaneous emphysema	Air escapes into tissue under the skin (as opposed to pneumothorax) after trauma, causing swelling and crackling sensation. (DISTINCT FROM EMPHYSEMA)
Other	Diseases and injuries not categorized above.
Healthy	No abnormal findings.



Google Colab

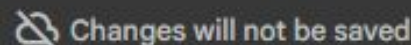
Google Colab

- Once we are done annotating, I will download the data and run in through a script that strips away the report texts (we can't let Google see them)
- I will then upload the anonymized data onto GitHub.
- The Colab will automatically download the data from there, no effort on your part required.
- To get the Colab notebook, go here:



https://colab.research.google.com/drive/1FGinFepX67gxicsL9O_V8S680CX1BhtD?usp=sharing

- The notebook has reader-only sharing; you can edit but not save changes, so don't worry about ruining anything
- If you want to copy it to experiment later, you can just click on



Changes will not be saved

<https://github.com/iml-r/heidelberg-nmt>

Google Colab

- If you're not familiar with Colab:
- It's a service by Google which allows you to run Python code blocks alongside text and visualizations
 - It also lets you use some very fancy hardware, but we won't be needing that today
- The Colab we will be using today has instructions written by me and designated empty spaces in some of the code blocks for you to solve the exercises

- If you run into any trouble with Label Studio, Colab, can't find a link, or don't understand something, just remember the rules:

Final note

- Rule 0:
 - **You are not alone. I'm here to help.**
 - **Don't panic!**
- Rule 1:
 - Everything (slides, data, links, explanations) can be found here:
 - <https://github.com/iml-r/heidelberg-nmt>

Label Studio – label set

Condition

Description

Pneumothorax	Air leaks into the space around the lungs, possibly collapsing part or all of a lung.
Fracture of unspecified body region	A bone break is detected, but the location is not listed in the other labels (e.g. facial fracture)
Cardiomegaly	Enlargement of the heart, often linked to heart disease or chronic high blood pressure.
Fluid overload	Excess fluid in the body, often affecting the lungs, due to kidney or heart failure.
Pleural effusion	Fluid buildup in the space between the lungs and chest wall.
Low bone mass disorders	Conditions like osteoporosis that weaken bones, increasing fracture risk.
Emphysema	Chronic lung damage where air sacs are destroyed, making it hard to breathe.
Traumatic subcutaneous emphysema	Air escapes into tissue under the skin (as opposed to pneumothorax) after trauma, causing swelling and crackling sensation. (DISTINCT FROM EMPHYSEMA)
Other	Diseases and injuries not categorized above.
Healthy	No abnormal findings.