Supplementary Materials

RadGraph: Extracting Clinical Entities and Relations from Radiology Reports

Saahil Jain[*]
Stanford University
saahil.jain@cs.stanford.edu

Ashwin Agrawal[*]
Stanford University
aag@cs.stanford.edu

Adriel Saporta[*]
Stanford University
adriel@cs.stanford.edu

Steven QH Truong
VinBrain, VinUniversity

Du Nguyen Duong
VinBrain

Tan Bui
VinBrain

Pierre Chambon
Stanford University

Yuhao Zhang
Stanford University

Matthew P. Lungren
Stanford University

Andrew Y. Ng
Stanford University

Curtis P. Langlotz[†]
Stanford University
langlotz@stanford.edu

Pranav Rajpurkar[†]
Stanford University
pranavsr@cs.stanford.edu

[*]Equal Contribution
[†]Equal Contribution

Preprint. Under review.

1 Additional tables

We report the top entities (Table A1) and relations (Table A2) in our train set after singularizing each entity.

Table A1: Top entities in train set

Anatomy		*Observation: definitely present*		*Observation: uncertain*		*Observation: definitely absent*	
Entity	%	Entity	%	Entity	%	Entity	%
right*	8.3	effusion*	4.3	atelectasis*	9.7	pneumothorax*	18.8
pleural*	8.0	normal*	3.9	effusion*	7.7	effusion*	14.0
left*	7.3	unchanged*	3.8	pneumonia*	7.3	consolidation*	8.1
lung*	6.4	opacity*	3.4	infection	4.9	acute	7.8
pulmonary*	5.2	tube*	3.0	edema*	4.6	focal	7.3
silhouette*	2.7	atelectasis*	3.0	small*	4.4	pneumonia*	4.6
heart*	2.4	mild*	2.8	aspiration	3.1	edema*	4.5
size*	2.4	small*	2.6	mild*	2.9	process	3.7
cardiac	2.2	moderate	2.4	consolidation*	2.4	abnormality	3.3
lower	2.2	stable	2.3	infectious	1.9	opacity*	1.4
mediastinal	2.2	clear	2.0	pneumothorax*	1.5	new	1.4
contour	2.1	edema*	1.8	hemorrhage	1.4	change	1.4
bilateral	2.0	tip	1.6	underlying	1.4	tube*	1.4
lobe	1.8	cardiomegaly	1.4	superimposed	1.2	large	1.3
volume	1.8	increased	1.2	chronic	1.2	congestion	1.2
vascular	1.6	low	1.1	scarring	1.2	fracture	0.9
cardiomediastinal	1.6	line	1.0	process	1.0	overt	0.8
upper	1.5	congestion	1.0	fluid	1.0	air	0.8
hilar	1.3	new	1.0	focal	1.0	free	0.7
chest	1.1	pneumothorax*	1.0	early	0.9	displaced	0.6

** Top 20 entities across all types.*

Table A2: Top relation pairs in train set

modify			*located at*			*suggestive of*		
Subject	Object	%	Subject	Object	%	Subject	Object	%
small*	effusion*	1.7	effusion*	pleural*	12.2	opacity	atelectasis	11.6
size*	heart*	1.7	edema*	pulmonary*	4.3	opacity	pneumonia	7.2
silhouette*	cardiac*	1.5	clear*	lung*	3.0	opacity	infection	4.6
focal*	consolidation*	1.5	normal*	heart*	2.2	consolidation	pneumonia	4.6
right*	pleural*	1.3	congestion*	vascular*	2.1	opacity	aspiration	2.8
volume*	lung*	1.3	low	lung	1.5	opacity	edema	1.9
left*	pleural*	1.2	normal	cardiomediastinal	1.4	blunting	effusion	1.6
lower*	lobe*	1.2	opacity	lung	1.3	marking	edema	1.6
bilateral*	pleural*	1.1	process	cardiopulmonary	1.3	low	crowding	1.4
contour*	mediastinal*	1.1	atelectasis	bibasilar	1.2	consolidation	atelectasis	1.2
left*	lung*	1.0	normal	hilar	1.1	opacity	effusion	1.2
moderate*	effusion*	1.0	normal	mediastinal	1.1	opacification	effusion	0.9
contour*	hilar*	1.0	fracture	rib	1.0	opacity	scarring	0.9
right*	lung*	0.9	consolidation	lung	1.0	opacity	consolidation	0.9
silhouette*	cardiomediastinal*	0.9	tube	chest	0.9	opacification	fluid	0.9
left	lobe	0.9	abnormality	osseous	0.9	opacity	hemorrhage	0.9
acute	process	0.8	pneumothorax	apical	0.9	consolidation	aspiration	0.7
mild	edema	0.8	opacity	lobe	0.8	opacity	fluid	0.7
right	lobe	0.8	normal	cardiac	0.8	enlarged	hypertension	0.5
tip	tube	0.7	edema	interstitial	0.8	prominence	hypertension	0.5

** Top 20 relations across all types.*

2 Annotation instructions given to radiologists

For this task, please label the reports assigned to you on our labeling platform (datasaur.ai). We expect each report to take approximately 10 - 25 minutes to label.

The labeling task is to find and label entities (**Anatomy** and **Observation**) in a given report as well as draw relations (**Suggestive Of**, **Located At**, and **Modify**) between entities.

Entities

An entity is a continuous span of text that corresponds to a physical entity of interest in radiology. For the purpose of annotation, we define the following two major types of entities:

- **Anatomy**: an anatomical body part that occurs in the radiology report. Examples of anatomy entities include *lung*, *left lower lobe of the lung* (multiple entities), or *aortic arch* (multiple entities).
  - **Anatomy vs. Anatomy Modifier**: in this schema, all anatomy modifiers are annotated as anatomy entities. In the case that an Anatomy modifies the scope or degree of a second Anatomy, a modify relation is added to denote the modification relationship between the two entities. See below for a definition and example of the modify relation. So, left lung would be two anatomy entities, where *left* modifies *lung*.
- **Observation**: an observation made from the images and associated with visual features, identifiable pathophysiologic processes or diagnostic disease classifications. Examples of multiple observation entities include *airspace opacity*, *mass*, *bilateral pleural effusion* or *pneumonia*. Observations can also include more general phrases like *increased*.
  - **Observation vs. Observation Modifier**: in this schema, all observation modifiers are annotated as observation entities. In the case that an Observation modifies the scope or degree of a second Observation, a modify relation is added to denote the modification relationship between the two entities. See below for a definition and example of the modify relation.
- Note that each entity is associated with an uncertainty attribute measuring the uncertainty level of an observation entity or anatomy entity. Each uncertainty attribute can have one of three values: **definitely present**, **uncertain**, or **definitely absent**.
  - When you label an anatomy, you will pick from the following options: *Anatomy::definitely present*, *Anatomy::uncertain*, or *Anatomy::definitely absent*.
  - When you label an Observation, you will pick from the following options: *Observation::definitely present*, *Observation::uncertain*, or *Observation::definitely absent*.
  - In some cases text spans for Anatomy or Observations are not continuous. In that case, each span should be labeled as Anatomy/Observation with a modify relation between them.

Relations

Relations are directed edges from one entity to another that are used to describe a relationship between two entities. We define the following three types of relations, in the form of "Relation Type (Entity Type, Entity Type)":

- **Suggestive Of** (Observation, Observation): a relation between two observation entities indicating that the status of the second Observation is inferred from that of the first Observation.
- **Located At** (Observation, Anatomy): a relation between an observation and an anatomy entity indicating that the Observation is related to the Anatomy. While Located At often refers to location, it can also be used to describe other relations between an Observation and an Anatomy. For example, in the sentence, *heart is normal*, *normal* is an Observation, *heart* is an anatomy, and Located At is a relation from *normal* to *heart*.
- **Modify** (Observation, Observation) or (Anatomy, Anatomy): a relation between two observation entities indicating that the first Observation modifies the scope of or quantifies the degree of the second Observation. This relation is often added when the first Observation is

an observation modifier that modifies the second Observation. The same logic applies to an anatomy entity modifying another anatomy entity.

Suggested Annotation Process

- Here is a suggested process for each sentence or sentence group:
  - Find the main anatomy (if present), and label as Anatomy.
  - Find any anatomic modifiers. Label these modifiers as Anatomy and link each to main anatomy using Modify.
  - Find all observations and label them as Observations.
  - Link all observations to the appropriate Anatomy using Located At relation.
  - Find any observation modifiers and label as Observations with modifier link to Observations they modify.
- Here is how this process would be applied to the following example: "The heart is top normal in size, though this is stable."
  - Main Anatomy is *heart*.
  - *size* is an attribute of *heart*, so it is also labeled as Anatomy with modify relation to *heart*.
  - *normal* and *stable* are Observations with a Located At relation to the main Anatomy, *heart*. Note that *top* is an Observation with a Modify relation to *normal*.

Other Specific Annotation Rules

- The main Anatomy may be an adjective or may be part of a compound term. For *pleural effusion*, *pleural* is Anatomy and *effusion* is Observation. A common discrepancy was labeling the entire phrase as the Observation.
- Words such as *size, volume, wall of, silhouette, structure, length, near, below,* are typically attributes or modifiers of Anatomy: e.g., *lung volume, heart size, cardiac silhouette, osseous structure, below the diaphragm, near the apex*.
- By convention, compound words, such as *cardiomegaly*, which means an enlarged heart and includes both Anatomy and Observation semantics, should be labeled as an Observation. In the same way, *pneumothorax* or *cirrhosis* are Observations with the anatomic location only implied by the term itself.
- Terms like *left*, *right*, and *bilateral* should be considered anatomy modifiers and labeled as Anatomy. For example, in *left kidney*, both *left* and *kidney* are Anatomy. For *bilateral pleural effusions*, *bilateral* is a modifier for the primary Anatomy (*pleural*).
- When an Anatomy and the modifier(s) are next to each other, identify a primary term and the modifier(s) separately for consistency even if they are adjacent. For example, *bilateral pleural*, should be labeled as two observations, with *bilateral* modifying *pleural*. This creates consistency with the case where the two tokens are distant in text, such as *bilateral small effusions* or *large effusions which are now bilateral*.
- A word in a sentence can sometimes modify a word in a previous sentence. For example, in the sentences, "The tube extends to the stomach. The tip is near the GE junction,", the word *tip* should have a modifier link to the word *tube* in the first sentence.
- Whenever there is any degree of uncertainty, mark the Anatomy or Observation as uncertain. For the phrases, *No signs of pneumonia* and *No evidence of pneumonia*, *pneumonia* is definitely absent. For the phrases, *No definite signs of pneumonia*, *No clear signs of pneumonia*, and *No obvious signs of pneumonia*, *pneumonia* is uncertain due to the additional qualifier.