

ShARe/CLEF eHealth 2013 Shared Task:
Guidelines for the Annotation of Disorders in Clinical Notes

Authors:

Noémie Elhadad, Columbia U
Wendy Chapman, U of California, San Diego & VA San Diego Healthcare System
Guergana Savova, Children's Hospital Boston & Harvard Medical School

Acknowledgements:

The Shared Annotated Resources (ShARe) project is funded by the United States National Institutes of Health (R01 GM090187).

The guidelines annotation and schema for the SHARe/CLEF eHealth 2013 shared task are available to the research community. When using the annotation guidelines in your own research and publications, please cite:

The ShARe Schema for the Syntactic and Semantic Annotation of Clinical Texts. Noémie Elhadad, Wendy Chapman, Tim O’Gorman, Martha Palmer, Guergana Savova. Under Review.

Table of Contents

ShARe/CLEF eHealth 2013 Shared Task:.....	1
Guidelines for the Annotation of Disorders in Clinical Notes	1
Table of Contents	2
Overview	3
1. Identifying a Mention of a Disorder (span and CUI)	3
1.1. A disorder mention is a concept in SNOMED-CT part of the Disorder Semantic Group.....	3
1.2 A disorder mention is a reasonable synonym of the lexical form of a SNOMED-CT Disorder Group CUI.....	4
1.3 A disorder mention is explicit	5
1.4 A disorder mention is the most specific disorder conveyed in the text	5
1.5 A disorder mention does not contain any negation in its span	6
1.6 A disorder mention does not contain any mention of temporality in its span	6
1.7 A disorder mention is not linked to any syntactic construct	7
1.8 The span for a disorder can be disjoint.....	7
1.9 A disorder mention is annotated even if it does not pertain to a patient.....	7
1.10 All the disorder mentions in a note must be annotated	8
2. Selecting the Appropriate CUI for a Disorder Mention	9
2.1 There should always be one CUI per mentioned disorder	9
2.2 There should be only one annotation per mentioned disorder	9
2.3 Use the context in which the disorder is mentioned to select the most appropriate CUI	9

Overview

The goal of this annotation task is to go through clinical notes, find mentions of disorders and annotate them.

Annotating **a disorder** consists of several steps:

- (1) **identifying a span of text** in the note that corresponds to the mention of the disorder and **mapping it to a CUI** (Concept Unique Identifier) from the provided terminology (SNOMED CT). The choice for the span of text and the CUI corresponding the disorder are tightly coupled, as the decision for one affects the decision for the other.

The remainder of this document goes through guidelines for annotating mentions of disorders.

1. Identifying a Mention of a Disorder (span and CUI)

1.1. A disorder mention is a concept in SNOMED-CT part of the Disorder Semantic Group

We define a **disorder mention** if as any **span of text** that can be **mapped to a concept** in the **SNOMED-CT** terminology, which belongs to the **Disorder semantic group**. A concept is in the Disorder semantic group if it belongs to one of the following UMLS semantic types:

- Congenital Abnormality
- Acquired Abnormality
- Injury or Poisoning
- Pathologic Function
- Disease or Syndrome
- Mental or Behavioral Dysfunction
- Cell or Molecular Dysfunction
- Experimental Model of Disease
- Anatomical Abnormality
- Neoplastic Process
- Signs and Symptoms

Please be aware that your notion of what a disorder is may be quite different than what is allowed by this set of semantic types. Be careful not to let your prejudice about what you consider a disorder get in the way of annotating disorder mentions.

Note that this definition of the Disorder semantic group differs slightly from the original definition from the UMLS, as we do not consider the Findings semantic type.

In our current annotation project, we rely on the **SNOMED-CT** terminology, part of the **UMLS version 2011AA**.

Example 1.1.a

This is a gentleman who presents with **colon cancer**.

The span “colon cancer” is a disorder mention. It corresponds to CUI C0007102 (Preferred term: Malignant tumor of colon) in SNOMED-CT, which is under the semantic type “Neoplastic Process.”

Example 1.1.b

No **chest pain**.

The span “chest pain” is a disorder mention. It corresponds to CUI C0008031 (Preferred term: Chest Pain) in SNOMED-CT, which is under the semantic type “Sign or Symptom.”

Example 1.1.c

The patient's glucose is 107.

There is no disorder mention in this sentence.

Example 1.1.d

The patient was continued on nebulizers while she was admitted.

There is no disorder mention in this sentence.

1.2 A disorder mention is a reasonable synonym of the lexical form of a SNOMED-CT Disorder Group CUI

A disorder mention must be annotated, even if the text span is not one of the lexical forms in the terminology, but is a **reasonable synonym**.

Example 1.2.a

Chief Complaint: left-sided **facial droop**

The span “facial droop” is a disorder mention. The lexical form is not part of SNOMED-CT, but can be mapped to CUI C0427055 (Preferred term: Facial Paresis) in SNOMED-CT.

Example 1.2.b

His **face** was **weak**.

The disjoint span “face ... weak” is a disorder mention. The lexical form is not part of SNOMED-CT, but it can be mapped to CUI C0427055 (Preferred term: Facial Paresis).

1.3 A disorder mention is explicit

A concept should not be annotated unless it is specifically mentioned by name. That is **descriptions of disorders should not be annotated**. Similarly, there should **not be any inference** made to decide that a disorder is mentioned.

Example 1.3.a

The patient's chem profile reveals a 138 sodium, 4.0 potassium, 106 chloride, 19 bicarb, 43 and 2.2 are the BUN and creatinine, indicating probable blood in the GI tract.

There is no disorder mention in this sentence, even though it describes the disorder GI bleed.

Example 1.3.b

EF 10-15%.

There is no disorder mention in this sentence, even though an ejection fraction of 10-15% strongly indicates the disorder congestive heart failure.

Example 1.3.d

She had a Crohn flare with symptoms of bowel obstruction that typically resolves with rehydration. Her current symptoms of reminiscent of this.

The first sentence contains two disorder mentions ("Crohn" and "bowel obstruction"). There is no disorder mention in the second sentence: "this," while referring to a disorder, is not a disorder mention.

1.4 A disorder mention is the most specific disorder conveyed in the text

Only the **most specific named disorder** may be annotated.

Example 1.4.a

The patient has a small bowel obstruction.

The sentence contains only one disorder mention, "small bowel obstruction." It corresponds to CUI Co235329 (preferred term: Small bowel obstruction). Only one mention is annotated, even though there are sub-spans of text corresponding to other disorders: the span "obstruction" can be mapped to CUI Co028778 (preferred term: Obstruction), and the span "bowel obstruction" can be mapped to CUI Co021843 (preferred term: Intestinal Obstruction). Because "small bowel obstruction" is more specific than "bowel obstruction" and "obstruction," it is the disorder chosen for this annotation.

Example 1.4.b

The patient was found to have left lower extremity DVT.

The sentence contains only one disorder mention, "lower extremity DVT." It corresponds to CUI Co340708 (preferred term: Deep vein thrombosis of lower limb).

The span “DVT” can be mapped to CUI Co149871 (preferred term: Deep Vein Thrombosis) is present, but is not annotated because it is part of a more specific disorder in the sentence.

Example 1.4.c

The patient has severe pre-eclampsia.

The sentence contains only one disorder mention, “severe pre-eclampsia.” It corresponds to CUI Co341950 (preferred term: Severe pre-eclampsia). The sub-span “pre-eclampsia” can be mapped to CUI Co032914 (preferred term: Pre-eclampsia) but is not annotated as it is more general.

Example 1.4.d

The patient has chronic gingivitis.

The sentence contains only one disorder mention, “chronic gingivitis.” It corresponds to CUI Co008684 (preferred term: chronic gingivitis). The sub-span “gingivitis” can be mapped to CUI Co017574 (preferred term: gingivitis) but is not annotated as it is more general.

1.5 A disorder mention does not contain any negation in its span

When looking for a concept in the terminology, do not map to concepts that have negations in them.

Reason: The negation will be annotated as a modifier of the disorder.

Example 1.5.a

No pain.

The span “pain” is a disorder mention (Co030193, preferred term: Pain), even though there is a SNOMED-CT CUI for the span “no pain” (Co234225, preferred term: Absence of pain).

1.6 A disorder mention does not contain any mention of temporality in its span

When looking for a concept in the terminology, do not map to concepts that have information about past or history.

Example 1.6.a

Hx of stroke.

The span “stroke” is a disorder mention (Co038454, preferred term: Cerebrovascular accident), even though there is a SNOMED-CT Disorder CUI for the span “Hx of stroke” (Co559159, preferred term: “H/O: CVA”).

1.7 A disorder mention is not linked to any syntactic construct

Most of the time, a disorder mention is either a whole or a subset of a noun phrase, including any prepositional phrase attached to a noun phrase. But in some cases, the disorder mention is not conveyed through a noun phrase only.

Example 1.7.a

The patient had a **tumor of the skin** removed.

The span “tumor of the skin” is a disorder mention. It corresponds to CUI Co037286 (preferred term: Skin Neoplasms). The determiner at the edge of the span “a” is not part of the annotation, but the preposition “of” is part of the disorder mention.

Example 1.7.b

A **tumor** was found in the left **ovary**.

The disjoint span “tumor ... ovary” is a disorder mention. It corresponds to Co919267 (preferred term: ovarian neoplasm).

1.8 The span for a disorder can be disjoint

When necessary, the span for a disorder can be disjoint. In such cases, it is important not to annotate the whole sentence as disorder span, but rather annotate only the keywords that describe the most specific concept.

Example 1.8.a

A **tumor** was found in the left **ovary**.

The disjoint span “tumor ... ovary” is a disorder mention. It corresponds to Co919267 (preferred term: ovarian neoplasm).

1.9 A disorder mention is annotated even if it does not pertain to a patient

Any mention of a disorder should be annotated in a clinical note. The fact that a particular disorder mention does not pertain to the patient is irrelevant.

Reason: The goal is to train and test NLP systems that identify any disorder mention. This is different from a retrieval system, which would do some inference about whether the presence of a disorder mention signifies that the patient has the particular disorder.

Example 1.9.a

The patient’s son has **schizophrenia**.

The span “schizophrenia” is a disorder mention (Co036341, preferred term: Schizophrenia). The fact that the patient’s son has the disorder will later be coded in one of the disorder’s modifiers.

Example 1.9.b

He should return to the ED immediately if any rash occurs.

The span “rash” is a disorder mention (C0015230, preferred term: Exanthema), even though the patient may never have any, does not have it currently, and may never have any. The fact that this is a conditional disorder is coded will be coded as one of the disorder’s modifiers.

Example 1.9.c

The patient was referred to the lupus clinic.

The span “lupus” is a disorder mention (C0409974, preferred term: Lupus Erythematosus), even though the patient may or may not have lupus. The fact that this is part of an institution will be coded in one of the disorder’s modifier.

1.10 All the disorder mentions in a note must be annotated

Every disorder mention must be annotated, no matter (i) what section of the note it appears in and (ii) whether a previous mention of the same disorder was already mentioned.

2. Selecting the Appropriate CUI for a Disorder Mention

Once a span for a disorder mention is identified, there still might be several CUIs available for that particular span. This section guides you through selecting the appropriate CUI for a disorder mention.

2.1 There should always be one CUI per mentioned disorder

A CUI is required for every disorder mention. If it is impossible to find a CUI corresponding to the disorder, **one should leave the slot empty as a last resort solution.**

Example 2.1.a

Atrial ectopy

There is a disorder in the UMLS, but not in SNOMED-CT (similarly for ectopy). This should be annotated as a disorder, but the CUI is left empty.

Example 2.1.b

Thick bronchial secretions

While “bronchial secretions” is a body substance, “thick bronchial secretions” can be interpreted as a disorder. There is a no disorder in SNOMED-CT for “thick bronchial secretions”. This should be annotated as a disorder, but the CUI is left empty.

Example 2.1.c

The patient was admitted with low blood pressure.

The span “low blood pressure” is a Finding in SNOMED-CT, and as such does not belong to our definition of the Disorder semantic group. In this case, however, because it does indeed describe a disorder, it should be annotated. The CUI is left empty.

2.2 There should be only one annotation per mentioned disorder

In many circumstances there will be a mentioned disorder, which is difficult to map exactly to a single CUI. It may be tempting to annotate the same disorder with multiple CUIs. However, for this annotation task, you may only annotate a mentioned disorder with a single annotation span and one corresponding CUI. The challenge is to choose the best CUI for the mentioned disorder along with the text span that best corresponds to that CUI.

2.3 Use the context in which the disorder is mentioned to select the most appropriate CUI

Whenever possible, use the context of the note to identify the most appropriate CUI for a given disorder.

Example 2.3.a

The patient complained of watering of the eye.

There are two candidate CUIs: C0152227 (Excessive tearing) and C2233621 (Watering or discharge from eye). Both are from the same semantic class “Sign or Symptom”. The definitions can help identify the most likely CUI for this particular note.

Example 2.3.b

The patient complained of numbness.

There are two candidate CUIs: C0028643 (numbness) and C0235018 (localized numbness). Both are from the same semantic class “Sign or Symptom”. In this case, the preferred term can help decide the most likely CUI for this particular sentence (i.e., the first one).