

Extended Guidelines for Symptoms Detection and Symptoms Relevancy Assessment in Medical Question Answering Cases

Benjamin Molinet, Santiago Marro and Maya Medjad

July 2022

1 Introduction

In this document, we specify the annotation guidelines we wrote for the symptom relevancy assessment task in Natural Language Processing (NLP). The task aims at predicting if a given symptom, together with some external knowledge available, is a relevant symptom for a given disease. The main objective of this task is to take a first step in explanatory argument generation by training models to i) detect the symptoms of a clinical case, ii) predict which of these symptoms are relevant for the target disease, and iii) generate explanations.

We wrote these annotation guidelines to build a corpus of medical Question Answering clinical cases based on the MEDQA dataset [1]. First, we extracted English cases from the MEDQA dataset, then we selected the data that describes clinical cases with questions asking for the most probable diagnosis (i.e., for each question about a disease, the possible diagnoses are provided as answers). We annotate different components with labels and definitions taken from the UMLS semantic network¹. The corpus has been pre-processed using the QuickUMLS[2], a tool that automatically detects and tags some components as “[Finding](#)” and “[Sign_or_Symptom](#)”.

In the remainder of the document, we specify the annotation guidelines we followed to build this dataset, including a definition and some examples of each annotated component.

2 Symptoms annotation

We keep the same definition as UMLS and adapt it for the [Sign_or_Symptom](#) component described as *an observable disease or condition in the clinical case including the symptoms from the past, the symptoms not related to the patient (family antecedents) and considering all kinds of diseases (physical, mental, ...)*.

¹<https://uts.nlm.nih.gov/uts/umls/semantic-network/T100>

Example 1 *He also has a 1-year history of joint and muscle pain in his calves and a 1-month history of intermittent, diffuse abdominal pain.*

Example 2 *She has a 10-year history of osteoporosis.*

Example 3 *His sister has anemia associated with recurrent nosebleeds and heavy menstrual periods; she also had postpartum hemorrhage that required transfusion.*

Examples 1, 2, and 3 show three examples of sentences we found in our MEDQA dataset where symptoms have been identified.

Example 4 *She says that despite the test results, she has had anxiety, insomnia, and a preoccupation with cancer since noticing the lump*

Example 5 *A 45-year-old woman has a 2-week history of increased anxiety, abdominal discomfort, irritability, and difficulty concentrating; she was robbed at knifepoint in a parking lot 3 weeks ago.*

Example 6 *Physical examination shows a 6-mm, red, ulcerated lesion with heaped borders.*

The identified components (i.e., symptoms) have to be as small as possible and we need to split them when multiple symptoms appear in a single sentence. As shown in Example 4 and 5, symptoms are split in multiple components and do not include the punctuation in the components boundaries. Symptoms should be self-contained as in Example 1 *joint and muscle pain*. They cannot be divided in *joint pain* and *muscle pain* as they constitute one single component. Example 6 follows the same idea.

Example 7 *She had an extensive abdominal operation 5 years ago for Crohn disease.*

Diseases are also annotated as symptoms like in Example 7 since the previous disease is likely to be a symptom of the current disease.

Example 8 *A 37-year-old woman, gravida 5, para 4, at 34 weeks' gestation comes to the emergency department because of vaginal bleeding for 2 hours.*

Example 9 *Five minutes later, there is bleeding from the nose, mouth, and intravenous sites.*

The boundaries of the identified components may include the body part affected by the symptoms when they are strongly related as in Example 8. In Example 9, the symptom *bleeding* is related to multiple body parts, so we annotate the locations separately as described in the next section.

Example 10 *On arrival, he is pale and in moderate respiratory distress.*

Example 11 *An 87-year-old woman, gravida 6, para 6, is brought to the emergency department because of a 24-hour history of severe **abdominal pain** and an **inability to void**.*

Example 12 *Abdominal examination shows mild diffuse **tenderness**.*

We omit quantifiers like *moderate* or *severe* and descriptive adjectives like *diffuse* as in Example 10, 11 and 12.

Following the labels of the UMLS semantic network, we started annotating the absence of symptoms with the “**No_Symptom_Occurence**” label that follows the exact same rules as regular symptoms with the only difference that this symptom is **not observed** in the clinical case or is resolved.

Example 13 *He has not had **chest pain** or **shortness of breath**.*

Example 14 *On cardiac examination, no **murmurs** or **gallops** are heard.*

Example 15 *Her **dysuria** has resolved.*

We do not include the negation in the annotation as shown in Example 13 and 14. Example 15 is the case of a resolved symptom which we do not consider relevant, therefore we annotate it as “**No_Symptom_Occurence**”.

In addition to the UMLS semantic network labels, we introduce a new annotation label called “**Relevant_symptoms**”. We annotate with this label every symptoms that appear for a given disease, according to previously retrieved knowledge. In order to detect a maximum of symptoms and relevant symptoms, we annotate the dataset in two steps: i) we find the “**Sign_or_Symptom**” components, and ii) for each annotated component, we manually check its relevancy according to our database and change the label. The database we use for this verification step is made up of the merged knowledge base from 3 sources, as highlighted in Table 1.

	DBPedia	WikiDoc	UMLS
Symptoms	X	X	X
Definitions	X		X
Concepts	X		
Raw text		X	

Table 1: Sources repartition of database data.

3 Extended annotation

Since a decision on a diagnosis cannot be based on symptoms alone, we decided to annotate more types of data that appear frequently in our clinical cases.

Example 16 *Pulse is 60/min, respirations are 18/min, and blood pressure is 100/62 mm Hg.*

Example 17 *During a 3-day period, his pulse increases from 82/min to 125/min, and blood pressure decreases from 124/72 mm Hg to 100/55 mm Hg.*

Example 18 *His immunizations are up-to-date.*

Example 19 *Test of the stool for occult blood is negative.*

Starting with the most recurrent components that appear in clinical cases, we annotate with the UMLS label “**Finding**” all the patient’s information given as measurements, evolution or test results (Examples 16, 17, 18 and 19). Same as with symptoms’ boundaries, findings should be as self-contained as possible like in Example 16.

Additionally, we annotate the absence of findings with the same rule as for symptoms’ components with “**No_Finding_Occurence**”.

Example 20 *Her pulse and blood pressure cannot be detected.*

Example 21 *Physical examination shows no abnormalities.*

When symptoms as *pulse* or *blood pressure* in Example 20 are not observed (negated), we choose to annotate them as no finding. We also consider important the fact that no other *abnormalities* are detected as in Example 21, discarding the lack of information.

We also decided to include the three following UMLS semantic network labels **Body Location or Region**, **Body Part**, **Organ**, or **Organ Component** and **Body Space or Junction**. In order to not misclassify the location labels, we choose to group them under the “**Body Location**” annotation tag.

Example 22 *Examination of the right upper extremity shows erythema and moderate edema.*

Example 23 *She has a mass in the right lower quadrant and hyperactive bowel sounds.*

Example 24 *A 4-year-old girl has the sudden onset of abdominal pain and vomiting.*

Example 22 and 23 differentiate the symptom from the body part where it occurs, while in Example 24 the symptom *pain* is strongly related to the associated body part, i.e., *abdominal*, therefore we annotate both of them as a symptom. As a criterion to annotated coherently the symptoms, as assumed that if the symptom semantics changes when removing the location, then we include the location in the component.

Clinical cases concerning diseases often refer to the past of the patient or mention disease’s duration. Therefore, we decide to annotate the “**Temporal_Concept**” from UMLS semantic network, including the duration, references to patient’s history or changes in symptoms.

Example 25 *For 8 weeks, a 52-year-old man with a 5-year history of type 2 diabetes mellitus has had deep burning pain in the ball of his right foot and big toe when the foot is raised above chest concentration.*

Example 26 *A 56-year-old man has had the painful weeping rash shown for 2 days.*

Example 27 *A 4-year-old girl has the sudden onset of abdominal pain and vomiting.*

The temporal concept component includes the time descriptors as *for 3 days* or *10 days ago* and the scales (*day, weeks, ...*) as in Example 25 and 26. We also consider the adverbs like *sudden* in Example 27 as temporal concept, because it can play a relevant role in the diagnosis.

Finally, given their main role in the diagnosis, we also annotated the gender and age components in our clinical cases with the “Age-Group” and “Population-Group” labels.

Example 28 *A 37-year-old woman comes to the physician because of shortness of breath for 3 months.*

Example 29 *A 16-year-old boy is admitted to the emergency department because of a knife wound to the left side of his chest.*

These labels are the easiest to find due to the document architecture, describing the clinical case by introducing the patient as shown in Example 28 and 29. The component also includes the *year-old* specification because of the possibility to encounter age measurement in months or weeks, especially for newborns.

4 Fully annotated examples

In this section, we report two examples that we annotated from the USMLE dataset. In these examples, the **symptoms** are written in orange, the **relevant symptoms** in bold, the **no occurrences** in red, the **findings** in blue, **locations** in green and **time descriptors** in violet. **Age** and **gender** are respectively in cyan and teal.

Example 30

A 37-year-old woman is brought to the emergency department because of intermittent chest pain for 3 days. The pain is worse with inspiration, and she feels she cannot take deep breaths. She has not had shortness of breath, palpitations, or nausea. She had an upper respiratory tract infection 10 days ago and took an over-the-counter cough suppressant and decongestant and acetaminophen. Her temperature is 37.2°C (98.9°F), pulse is 90/min, and blood pressure is 122/70

mm Hg. The lungs are clear to auscultation. S1 and S2 are normal. A **rub is heard during systole**. There is no **peripheral edema**. An ECG shows normal sinus rhythm and **diffuse, upwardly concave ST-segment elevation** and **PR-segment depression in leads II, III, and aVF**. Which of the following is the most likely diagnosis?

Example 31 A previously healthy **34-year-old woman** is brought to the physician because of **fever** and **headache for 1 week**. She has not been exposed to any **disease**. She takes no medications. Her **temperature is 39.3°C (102.8°F)**, **pulse is 104/min**, **respirations are 24/min**, and **blood pressure is 135/88 mm Hg**. She is **confused** and **oriented only to person**. Examination shows **jaundice of the skin and conjunctivae**. There are a few scattered **petechiae** over the **trunk** and **back**. There is no **lymphadenopathy**. Physical and neurologic examinations show no other **abnormalities**. **Test of the stool for occult blood is positive**. Laboratory studies show:

- Hematocrit 32% with fragmented and nucleated erythrocytes
- Leukocyte count 12,500/mm³
- Platelet count 20,000/mm³
- Prothrombin time 10 sec
- Partial thromboplastin time 30 sec
- Fibrin split products negative
- Serum
- Urea nitrogen 35 mg/dL
- Creatinine 3.0 mg/dL
- Bilirubin
- Total 3.0 mg/dL
- Direct 0.5 mg/dL
- Lactate dehydrogenase 1000 U/L

Blood and urine cultures are negative. A CT scan of the head shows no **abnormalities**.

Which of the following is the most likely diagnosis?

5 Relations annotations

We also decided to annotate the relations between symptoms/findings and location/temporal concepts. We propose only one general relation called “**related**” from the symptom/finding *to* location/temporal concept.

Example 32 *There are a few scattered **petechiae** over the **trunk** and **back**.
With the annotated relations:*

- **petechiae** \Rightarrow **Related** \Rightarrow **trunk**
- **petechiae** \Rightarrow **Related** \Rightarrow **back**

Example 33 *A **37-year-old woman** is brought to the emergency department because of intermittent **chest pain** for **3 days**
With the annotated relations:*

- **chest pain** \Rightarrow **Related** \Rightarrow **for 3 days**

In Examples 32 and 33, we headed the relation from symptoms to location, and from symptoms to the temporal concept.

6 Edge cases

As for the symptoms, in some clinical cases we can encounter some enmeshed components as illustrated in Example 6. Here both *red* and *ulcerated* refer to the *lesion*, but they cannot be split in two self-contained symptoms’ component. The solution we choose is to include both components in one as in the example.

Example 34 *The **symptoms are moderately exacerbated by exertion**.*

Example 34 shows a symptom triggered by something in particular, this action is potentially relevant, so we include it in the annotated component.

Example 35 *He has had a **lot of energy for work** but often is **distracted** to the point that he does not complete assigned tasks. He frequently **stops working** on his own tasks to attempt to develop greater efficiency in his shop. He states that he is **delighted with his newfound energy** and reports that he now **needs only 4 hours of sleep nightly**.*

Example 35 shows a complicated clinical case where symptoms are put forward through common words as *stops working* or *lot of energy for work*. We consider these symptoms as “hard” to be detected automatically. Nevertheless, we annotated them.

Example 36 *The **left pupil is larger than the right and reacts sluggishly to light**.*

Example 36 represents a comparison between two body parts which are enmeshed, since we cannot divide them as *left pupil is larger than the right* and *left pupil reacts sluggishly to light*, we consider it as a single long symptom.

Example 37 *He has been your patient since early adolescence, and he has a history of truancy, shoplifting, and two attempts to run away from home. He dropped out of high school in his senior year. He was fired from his most recent job because he threatened a coworker with a hammer.*

Mental diseases as in Example 37 are quite hard to annotate relevant because of the peculiarity of their symptoms, so we follow and match as much as possible the symptoms available in the external knowledge database to annotate these examples.

Example 38 *After the seizure, she was confused and had difficulty thinking of some words.*

In some patient introduction, the gender is revealed later and using the gender pronouns as in example 38

References

- [1] Jin D, Pan E, Oufattole N, Weng WH, Fang H, Szolovits P. What disease does this patient have? a large-scale open domain question answering dataset from medical exams. *Applied Sciences*. 2021;11(14):6421.
- [2] Soldaini L, Goharian N. Quickumls: a fast, unsupervised approach for medical concept extraction. In: *MedIR workshop, sigir*; 2016. p. 1-4.