

NTCIR11 MedNLP2 Annotation Guidelines

Version 1.0
Last Modified: 2014/01/31

NTCIR MedNLP Organizers
mednlp.office@gmail.com

1. Introduction

This is the official NTCIR11 MedNLP2 Guideline for annotation. The samples in **BLUE** are examples that should be marked. The samples in **RED** are examples that should NOT be marked.

NTCIR11 MedNLP2 のアノテーション・ガイドラインです。青色でマークされているものはアノテートされる箇所、赤色でマークされているものはアノテートされない箇所です。

2. Material

This material contains two types of data: (1) Dummy Patients' Medical Reports (**D-Rep**), and (2) Questions from the past State Examinations (**Q-Rep**).

模擬病歴報告(以降, D-Rep)と国家試験問題(以降, Q-Rep)という2つの模擬文章から構成されます。

D-Rep.

This corpus is constructed from dummy medical reports that doctors have written for their 'dummy' patients.

模擬病歴報告とは、患者が退院する際や他の医師に紹介する際に、その患者の症例を理解することを目的として書かれる、実際の病歴報告文書を模して医師によって書かれたものです。

Q-Rep.

This corpus is from the actual past state examinations. The question parts and graphic parts are eliminated.

医師国家試験問題に臨床問題中の患者の状況を述べた文章です。

D-Rep. (MedNLP-1)	Q-Rep.
----------------------	--------

Disorder of the Alimentary Tract (消化器疾患)	4 (4)	8
Liver, Biliary Tract & Pancreas (肝・胆・膵疾患)	2 (2)	7
Cardiovascular System (心臓・脈管疾患)	10 (12)	7
Endocrinology, Metabolism & Nutrition (内分泌・代謝疾患)	7 (5)	6
Disorders of the Kidney & Urinary Tract (腎・泌尿器疾患)	2 (4)	6
Immune System & Immune-Mediated Injury (免疫・アレルギー性疾患・膠原病)	5 (5)	4
Disorders of the Hematopoietic System (血液・造血器疾患)	2 (1)	5
Infectious Disease (感染症)	4 (6)	6
Disorders of the Respiratory System (呼吸器・胸壁・縦隔疾患)	11 (11)	8
Total	47 (50)	57

3. Annotation of Date time

Date and time/tense related expressions should be marked as follows.

以下の方針により「日時」に関わる表記をマークします。

【compound noun】 Time related compound nouns should be marked.

If a compound noun contains terms that are not related to tense nor any other time related information, such terms should be excluded from the t-tag.

【複合名詞】 時間表現を示す名詞から構成される複合名詞をマークします。時間表現を示さない名詞が複合名詞に含まれる場合は、その部分を省いてマークします。

<t>計 1 4 日間</t>CAZ 投与した。
 <t>7 日間継続</t>の後、退院した。
 <t>2 月 8 日帰宅時</t>に異常を訴えた。
 治療<t>開始 2 ヶ月目</t>に結核菌培養が陰性となったため退院とした。
 ペンタミジンを<t>1 週間</t>投与。

【non numerical expression relating tense】 As a general rule, numerical expressions which indicate date and/or time should be marked. Non-numerical expressions should NOT be

marked, unless they indicate specific dates and/or time.

【数詞の入らない表現】原則として数詞の「時間」単位をマークします。数詞の入らない表現については、日付または時間のいずれかの特定が可能な場合にはマークします。

<t>4月8日</t>再度上部消化管内視鏡施行し、治癒傾向にあると判断し、<t>同日より</t>三分粥の潰瘍食開始。

【relative time】 Both ‘absolute’ and ‘relative’ time/tense expressions should be marked.

【相対時間】 相対時間もマークします。

発症から<t>3時間以内</t>である。

【cycle/frequency expression】 Expressions indicating cycles and/or frequencies should NOT be marked.

【頻度表現】 「回」や「毎」等、サイクルを示すものはマークしません。

月に約1回の割合で輸血を受けている。

【duration】 The whole clauses and phrases of time expression including punctuations and hyphenations (e.g. “—”, “～”) should be marked together.

【期間表現】 「から」や「より」「～」や「—」を用いた期間を表す時間表現にはまとめてマークします。

<t>2月20日より</t>プレドニゾロン45mg内服に切り替え、現在入院継続。
<t>H28年頃～</t>食事療法

【non clinical expression】 Expressions that are not directly related to medical information should NOT be marked.

【非臨床時間表現】 病状や治療に関連がない表現はマークしません。

40年来の専業農家。

【case marker】 Japanese case markers [Jo-shi], such as “の” should NOT be marked: Function words should NOT be included in t-tags.

【助詞】 助詞はマークしません。

<t>前日</t>の<t>27日昼</t>に退院した。

【life expectancy】 Patients’ life expectancies should NOT be included in t-tags.

【余命表現】 余命に関する時間表現はマークしません。

Stage 2となり、平均余命1.7年である。

4. Annotation of Symptom & Diagnosis

以下の方針に基づいて、病状、患者の症状、病名等に c-タグを付与します。
Symptoms & diagnoses related expressions should be marked as follows.

【compound noun】 Each noun compound word should be marked as the each noun phrase (concept) as a whole.

【複合名詞】 複合名詞をまとめてマークします。

<c>両側肺門リンパ節腫大</c>, <c>軽度網膜血管炎</c>

【verbal phrase】 Verbal phrases should NOT be marked.

【動詞句】 動詞句はマークしません。

両足が痛み右膝が腫れてきた
皮膚はいずれも硬くなった
夜はなかなか寝つけない, とのこと。

【body part and name of medical examination】 The expressions of body parts, the disease names included in the medical examinations, and/or the names of the pathogenic bacteria of the examinations should NOT be marked.

【検査名および身体部位】 身体部位あるいは検査, 及び, 検査対象の菌のみの表記にはマークしません。

生検蛍光抗体法で糸球体の免疫グロブリンは陰性であった。

【disease identification test】 If and only if the existence of the certain virus represents a single particular disease, the virus should be marked.

【疾患特定検査】 結核菌のように, 単なる検査名であっても, 菌が存在することで特定の病状を含意する場合はマークします。

喀痰 Gram 染色で, <c>Gram 陽性双球菌</c>の白血球による貧食像を多数認める。

【ICD contained examination】 Despite the cases of medical examination, they should be marked, if and only if the ICD codes are available.

【ICD 収載検査】 単なる検査名であっても, それ自体で ICD が存在する場合はマークします。

<c>ツベルクリン反応陽性</c>。

【non-disease based phrase】 The noun phrases that contain disease names as the parts of the medical tests and/or the surgeries should NOT be marked.

【疾患を構成しない名詞句】 検査や手術の名称に含まれる病名はマークしません。

左下肢静脈瘤手術を行った。
梅毒定性（-）。

【general description】 When the expression describes ‘general’ information about the disease and/or the name of the clinic, it should NOT be marked.

【一般的説明】 患者の症状でなく、一般的な病の説明として病名や施設名における病名はマークしません。

あたらしい分類では、自己免疫型の糖尿病は type 1 A と名付けられている。
かかりつけの■■■内科リウマチ科クリニックより当院紹介受診となった。

【non-alphabetical character】 Non-alphabetical characters and non-numerical characters (e.g. “↓”) should be marked with the previous noun phrases, if and only if the marks represent the conditions and are attached to the previous noun phrases to form the names of the disease.

【記号】 症例に“低下”という意味合いで“↓”がみられ、かつ名詞節と接続する場合、矢印も込みでマークします。

<c>皮膚ツルゴール↓</c>（皮膚ツルゴール低下は脱水症、特に低張性脱水症）

【exemplifications】 The expressions that describe the degrees or tendencies of disease conditions, such as “Teido 程度,” “Hodo ほど” and “Keikou 傾向,” should be marked. The expressions that describe instabilities and variables, such as “Tou 等” and “Nado など,” should NOT be marked.

【例示】 「程度」、「ほど」はマークする。「等」、「など」はマークしません。

<c>血痰程度</c>の軽度の症状であった。
体幹下肢の<c>皮膚病変痲皮化傾向</c>のため、メシル酸ガベキサートも終了となった。
<c icd="R53_">全身倦怠感</c>などの自覚症状が今回入院半年前から悪化してきた。

【modality related word】 The words and phrases that suggest modalities of the symptoms (e.g. positive 陽性/ negative 陰性/ prevention 予防/ deterioration 悪化/ emergence 出現/ decline, depression 低下/ enlarged 拡大/ increase, elevation 上昇/ normal 正常（正）/ dilate, enlarge 拡張/ diminution 縮小/ progression 憎悪/ change 変化/ decrease, fall 減少/ recurrence, relapse 再発/ continuum 継続/ anamnestic 既往/ ~able 可能) should NOT be included in c-tags.

【モダリティ関連語】 モダリティに影響する語（陽性/陰性/予防/悪化/出現/低下/拡大/上昇/正常（正）/拡張/縮小/憎悪/変化/減少/再発/継続/既往/可能）はマークしません。

<c modality="negation">ツ反応</c>陰性であった。
<c icd="B59_">ニューモシスチスPCR</c>陽性であった。
皮膚の<c icd="L299">発赤搔痒感</c>出現し、ペンタミジン終了とした。

Except, some modality-related words should be included, if and only if they are connected to compose the standard disease names.

ただし、モダリティに影響する語が他の語と結びついて病名とされている場合は、まとめてマ

ークします。

アミカシンの副作用である、`<c>腎機能異常</c>`は見られなかった。

【Separation】 Interpoints (・), slashes (/), hyphens (-), and/or any other punctuations should separate annotations: Separated phrases should be marked individually.

【セパレーション】 中黒 (・), スラッシュ (/), ハイフン (-) や、読点をはさむ場合は、べつべつにマークします。

明らかな`<c>出血</c>`、`<c>梗塞</c>`の所見はなし。

Except, when a punctuated phrase composes a single concept, it should be marked as a whole. ただし、まとまって1つの表現になっている場合、まとめてマークします。

`<c>ウイルス性、細菌性肺炎</c>`疑いあり。
`<c>腸蠕動音低下、亢進</c>`ともになし。

5. Modality Attribute

Patients' symptoms & diagnosis should be marked with the following 4 types of modalities. 症状と診断表現に関しては、以下の4種類のモダリティを属性としてマークします。

Positive	Actually recognized symptoms. (Default) その症状が実際に認められた場合
Negation	Symptoms that are NOT recognized. その症状が実際には認められなかった場合
Suspicion	Suspected and unconfirmed diseases. 病名についての推測や疑いの場合
Family	Diseases of the patients' family members. 家族歴としての病名の場合

37℃台の`<c>微熱</c>`とを主訴に来院した。
`<c modality="negation">圧痛</c>`や`<c modality="negation">発赤</c>`なし。
当初`<c modality="suspicion">両側丹毒</c>`が疑われた。
9人兄弟で姉2人が`<c modality="family">糖尿病</c>`。

In cases of the condition change for the better, they should be marked with "negation."
「改善」、「縮小」など、症状の好転を示す場合にも negation を用います。

`<c modality="negation">喀痰</c>`は改善した。

When diseases require two or more modalities, they should be marked by separating each

modality with commas (,).

モダリティを2種類以上伴う場合は、カンマで区切り併記します。

`<c modality="negation,family">大腸癌</c>の家族歴なし`

6. ICD Attribute

C-tags should include the ICD-10 codes.

MEDIS Standard Masters (ICD Taiou Hyoujyun-Byoumei Master), the application for corresponding ICD codes with standard disease names, can be downloaded from the MEDIS HP¹.

症状と診断表現に関しては、ICD-10 コードを属性として付与します。

MEDIS のページ¹から ICD 対応標準病名マスターがダウンロードできます。

ICD codes are normally consisted of single alphabet and 3 numerical digits. When the codes only have 2 or 1 digit(s), and/or when it is impossible to specify the codes from the contextual information, the last 1, 2 or 3 digits should be supplemented with 1, 2 or 3 underbar(s) (_) to equalize ICD codes to be in 1-alphabet and 3-digits-forms.

通常、ICD コードは3桁から構成されますが、テキストから十分な情報が得られず、コードを確定できない場合は、特定できない末尾の1~3桁にも、'_' (アンダーバー) を付与します。

I V H 刺入部に `<c icd="R21_">発赤</c>` を伴う `<c icd="A__">感染徴候</c>` みられ中心静脈ルート抜去。

Similarly, when the ICD codes contain 4 numerical digits, the last 4th digit should be omitted to equalize every ICD code in 3-digits-form.

4桁のものについては、最後の1桁を省略し3桁とします。

`<c modality="negation" icd="M791">筋痛</c>` は改善し、さらにCRPは陰転化した。
* {`<c>筋痛</c>`の icd は M7919}

¹ <http://www2.medis.or.jp/stdcd/byomei/index.html>