Medical Synonym Annotation Guideline

Version 2.0 (2015-5-1)

Authors: Jinying Chen, Elaine Freund, Hong Yu

Contents

This annotation task is to judge whether the **candidate terms** found by an automatic system is a synonym, related term, or irrelevant term for a **target medical term**. The related annotation concepts are defined in Section 2 (Related Annotation Concepts).

1. Accessing Annotation Website

http://ummsres14.umassmed.edu/py/syneval/second

(Please use Firefox to access the tool. IE does not work properly with this tool. Other browsers may or may not work properly.)

The annotation webpage looks like the follows (see Fig 1).

Figure 1 Annotation User Interface (UI)

2. Related Annotation Concepts

Target term (marked by **red circle** in Fig 1): the term for which the automatic system finds candidate synonyms. A target term can be medical or non-medical term. The first step of the annotation is to filter out non-medical terms.

Medical term: Any terms (including abbreviations/acronyms) which often occur in medical reports, diagnostic notes, biomedical or medical papers (e.g., papers about drug safety, pathology of diseases, clinical trials, etc.) can be counted as medical terms.

Rules of Thumb/Hints:

(1) Diseases, symptoms, diagnosis, treatments, technologies and medical instruments for disease prevention/diagnosis/treatment and health preservation, and drugs as well as disease causing (pathogenic) bacteria/virus/chemicals/plants/animal etc. are typical medical terms

(2) If you are unsure about a term's meanings, you can search it in the Wikipedia and the online medical dictionaries listed in Appendix A. It is recommended to check at least two resources before **skipping** this term as "undecided".

(3) Animals and plants are generally treated as non-medical terms except for well-known species that have medical use or cause diseases. A quick reference is medilexicon and wikipedia

- http://www.medilexicon.com/medicaldictionary.php
- https://en.wikipedia.org/wiki/Special:Search?search=&go=Go

For example, searching "Ixodidae" in medilexicon shows the following definition.

"… It includes the genera Ixodes, Hyalomma, Amblyomma, Boophilus, Margaropus, Dermacentor, Haemaphysalis, and Rhipicephalus, species of which transmit many important human and animal diseases and cause tick paralysis; they occasionally attack humans, a few habitually so."

Some examples for non-medical terms: crotalus (generic venomous snake), cyanobacteria, drug nomenclature

(4) Small molecules or chemicals are generally treated as non-medical terms unless they are generic drug names, or have common medicinal usage in the strict sense, e.g., fluoxetine, glucosamine for arthritis.

(5) For terms with **multiple meanings (i.e., senses)**, it is treated as a medical term if only it has a meaning falling into this category. If a target term has multiple medical meanings, **use the dominant (most common) medical meaning** of the target term when making judgment on the relevancy of its candidate terms.

For example, the verb "indicate" has two meanings: (1) point out/show and (2) suggest as a desirable or necessary course of action. The second meaning has medical use, e.g., "The treatment is likely to be indicated in severely depressed patients". Using the second meaning of this verb when judging its candidate synonym terms.

See another example in .

Candidate term (marked by **green box** in Fig 1): the candidate synonym terms that the automatic system has found for the target medical term. A candidate term can be a synonym, a related term, or an irrelevant term of the target term.

Synonym: a synonym of a target medical term has exactly the same meaning as the term

Example 1: "high blood pressure" is a synonym of "hypertension"

Example 2: "Temperature Gradient Gel Electrophoresis" is a synonym of "TGGE"

Example 3: "true pelvis" is a synonym of "lesser pelvis"

Example 4: "heart attack" is a synonym of "myocardial infarction"

Related term: a term that is related to the target medical term in meanings:

- **hyponyms** (specific instances of general terms, e.g., "heart disease" is a hyponym of "disease") [1]
- **hypernyms** (general terms of specific instances, e.g., "disease" is a hypernym of "heart disease")
- words **derived** from the same root word and sharing similar meanings (e.g., noun and verb forms of the same root word, single and plural forms for the same noun)
- medical terms that have **high correlation** (e.g., a very typical symptom for a disease or a medical event that is typically caused by a disease)

Examples of Related Terms

Example 1: "allergic inflammation" is related to (a hypernym of) "allergic rhinitis"

Example 2: "Ixodidae" is related to (a hyponym of) "ticks"

Example 3: "tranquilizers" is related to (share the same root word and have similar meanings) "tranquilize"

Example 4: "heart disease" is related to (has high correlation with) "myocardial infarction"

---

[1] See http://en.wikipedia.org/wiki/Hyponymy_and_hypernymy for more background

(Note: "heart disease" is the typical cause for the medical event "myocardial infarction")

Example 5: "high blood sugar" is related to (has high correction with) "diabetes"
(Note: "high blood sugar" is a very typical symptom of "diabetes")

Rules of Thumb/Hints:

(1) A way to check the relevance of two terms is to search google (www.google.com) by keyword combination, e.g., "Ixodidae"+"hard ticks". The answer can often be found in the top-5 articles (the articles in the first page, e.g., Wikipedia article, etc.)

(2) For a given concept, synonyms should be of the same form, that is, both should be singular, plural, past, present, or future, etc. For example, "sleeping pill" and "tranquilizer" are synonyms, "sleeping pills" and "tranquilizers" are synonyms.

Different forms of the concept word or its synonyms are related terms. For example, if the concept is "tranquilizers" (plural), both "sleeping pill" (singular) and "transquilizer" (singular) are its related terms.

(3) Very general terms (e.g., deficiency, fracture and attack etc.) are not treated as medical terms. They gain medical meanings only in conjunction with another word (e.g., bone fracture). These general terms are not treated as related terms of their medical hyponyms (e.g., vitamin deficiency, bone fracture, heart attack etc.).

Irrelevant (rejected) term: a term which is not a synonym or related term for the target medical term and should be rejected

Example 1: "attack" is irrelevant to "heart attack"

Example 2: "integerrima" is irrelevant to "ilex integerrima"

Example 3: "infarction" vs. "heart attack"

Example 4: "high fever" vs. "flu"

(Note: "high fever" can be caused by many kinds of diseases in addition to "flu")

Example 5: "white blood cell count" vs. "white blood cell count"
(Note: reject a candidate if it is exactly the same as the target term)

Rules of Thumb/Hints:

(1) Derivative phrases from the target term is treated as an irrelevant (rejected) term if they don't have similar meanings to the target term or they carry significantly new meanings (e.g., "sedative effect" and "sedative")

3. Instructions for Annotation

Step 1: enter your **unique user name**

Note: Each time you launch the annotation tool, you are required to input your unique user name (FirstnameLastname, e.g., JohnSmith) in the first page. The user name will automatically show up for the rest pages.

Step 2: If the target medical term is not a real medical term (e.g., a person name or a country name such as "Republic of Malawi"), check the "Non-Medical" box (marked by the **blue circle** in Fig 1). **Then wait for about 2 seconds. The UI will automatically change to the next instance**.

If you are unsure whether this term is medical or non-medical after 3-minute thinking and research on the web, check the "Skip This Term" box (marked by the **orange circle** in Fig 1). **Then wait for about 2 seconds. The UI will automatically change to the next instance**.

If the target term is a medical term, go to Step 3.

Step 3: categorize each candidate term into one of the three categories (synonym, related, irrelevant/rejected terms) by dragging the term to the corresponding column

Step 4: click the "submission" button (blue button) to submit your annotation

Appendix A. Resources for Medical Terms/Abbreviations

Medical terms:

- MedicineNet: http://www.medicinenet.com/script/main/alphaidx.asp?p=a_dict (interactive UI for dictionary lookup)
- Medilexicon: http://www.medilexicon.com/medicaldictionary.php (interactive UI for dictionary lookup)
- Western Medicine: http://www.macmillandictionary.com/us/dictionary/american/western-medicine
- MedlinePlus: http://www.nlm.nih.gov/medlineplus/mplusdictionary.html

Medical abbreviations:

- http://www.medilexicon.com/medicalabbreviations.php (interactive UI, 200,000 terms)

Appendix B. Q&A Lists

1. How to treat misspellings?

- If the target term is misspelled, treat this term as "non-medical" term.
- If candidate term is misspelled, reject it.

2. How to treat terms that have punctuation markers, such as "high blood pressure."

- If the punctuation is the last character of the term (as the above example), treat it as correct.
- Otherwise, treat it as a misspelling and follows the rules in 1.

3. For medical terms with non-medical meanings, how to categorize the non-medical synonyms of these terms? For example,

Concept: emigration

Medical def: passage of white blood cells though the endothelium and wall of small blood vessels.

Some of the words the synonym tool provided:
Countries of origin
Emigrant
Arnaud Giroux
Leaving
Human flight …

Ans: In this case, reject these non-medical synonyms. Please refer to Section 2 **multiple meaning** for detailed instructions.

4. "Carotene" is the precursor to "vitamin-A", but they don't share the same word root nor literal meaning. Is "carotene" related to "vitamin-A"?

Ans: Reject "Carotene" for "vitamin-A". The definition of "related terms" in this task is narrow and strict. We current only focus on the cases described in Section 2.

5. Plants with medical use
        Concept: Salvia officinalis
        Candidates: Meramiyeh, Sage

"Salvia officinalis" is a medicinal plant in Lamiaceae family, salvia genus (from Wikipedia)

    a. Are medicinal plants and other things with medical uses but are sometimes not listed in common western medical dictionaries considered medical terms?

Ans: Yes, if there is evidence from other reliable resources (Wikipedia or other dictionaries).

    b. Assuming medicinal plants are medical terms. How to categorize plants/species that share the same genus or family?
For example, "Meramiyeh" is also from Lamiaceae family, are "Meramiyeh" and "Salvia officinalis" related?

Ans: No. The definition of "related terms" in this task is narrow and strict. We treat a target term's hypernyms and hyponyms (see Appendix A for definitions) as its related terms. However, we don't treat "apple" and "pear" as related terms although they are both "fruit".

    c. Also, Lamiaceae family may be referred as the "sage", so are "sage" and "Salvia officinalis" related?

Ans: Yes.

Appendix C. More Examples

Example 1. high blood glucose vs. high blood sugar vs. diabetes

From this definition from NIDDK

> "People with diabetes have high blood glucose, also called high blood sugar or hyperglycemia."
> high blood glucose = high blood sugar = hyperglycemia ("=" means "is a synonym")

High blood sugar is a typical symptom of Diabetes, but not uniquely related to Diabetes. That is, if a person has *high blood sugar*, it does not necessarily mean he/she has *diabetes*. So the two terms are not synonyms.

Example 2. Advocacy group

advocacy_organization is a synonym.
Advocacy_center, interest_organization, lobby_group and special_interest_group are hyponyms/related terms.

Example 3. table salt
Synonyms = salt, sodium_chloride
Related: salted
Reject: table, human_consumption, sodium_and_chloride

Example 4. neuroprotection
Related = neuroprotectant, neuroprotectants, neuroprotective, neuroprotector
Rejected= protective, neuroportective_drugs

Example 5. standardization, law enforcement, etc.
These terms are very common in other domains (e.g., business, politics etc.) as well. We can treat them as non-medical terms.