

How to use the web platform to annotate biomedical entities

First, click on the following website address (if it does not work, please, email us, asap): <https://promoting-connection-closing-raw.trycloudflare.com>

NOTE1: Every time, when you enter the annotation page, please, open this document and follow the address posted on this page.

NOTE2: Please, prioritize the annotation and flagging of historical entities and prioritize entity types: disease, symptom, anatomy, phenotype. You can check the details about the classes of the ontology that we use — here:

<https://bioportal.bioontology.org/ontologies/DOID?p=classes>

Prioritize the entities that are important for the further analysis of:

1. **Cause-based classification of disease** (Infectious, Genetic, Environmental, Nutritional Deficiency, Poisoning, Immune system disease, Metabolic disease)
2. **Classification based on the affected anatomical system** (cardiovascular, respiratory, digestive, nervous, reproductive, musculoskeletal)
3. **Symptom-based and Syndrome Classes** (Syndrome, Pain disorder, Inflammatory disease, fever-related, paralysis, etc.)
4. **Classifications related to Social constructs** (mental disorder, behavioural disorder, developmental disorder)

The page includes the following sections relevant to you:

The screenshot displays the 'Entity Linking Annotation Tool' interface. At the top, a dark header bar contains the tool's name and statistics: 'Total Samples: 466', 'Annotated: 8', 'Auto Entities: 437', and 'Manual Entities: 51'. The main workspace is divided into three primary areas. On the left, a text editor shows a medical document snippet with several terms highlighted in green, including 'cretinism', 'dwarf growth', 'goitrous', 'cretinous', 'goitre', 'thyroid gland', 'myxocdema', 'cachexia strumipirva', 'measles', and 'whooping cough'. A blue box labeled '1. Text' highlights the text area. In the center, a blue box labeled '2. Annotations' highlights the green-highlighted terms. On the right, a blue box labeled '3. Entity records' highlights a panel titled 'Detected Entities' which shows details for the entity 'goitre', including its ID (DOID:1459), namespace, status flag (set to 'None (default)'), and position in text (247 - 253). At the bottom, a blue box labeled '4. Navigate and Save Buttons' highlights a control bar with 'Previous', 'Next', and 'Save Current' buttons. Below this bar is a footer with a 'Sample Index' (7 of 466), a 'Go' button, a file path, and a 'Show Automatic Labels' checkbox.

Entity Linking Annotation Tool

Total Samples: 466 Annotated: 8 Auto Entities: 437 Manual Entities: 51

File: BMJ_1929_03_16_vol001_nr3558_art043_pmc2450088.txt
Line: 44

1. Text

R. [redacted] (arg., January, 1929, p. 371) defines **cretinism** as a condition characterized by **dwarf growth** associated with [redacted]. It occurs among the children of **goitrous** or **cretinous** parents in districts where **goitre** and cretinism are endemic, or sporadically in goitre free districts in the children of healthy [redacted] parents. Cretinism is a condition of the **thyroid gland** and is allied to the **myxocdema** of adults. There are three definite [redacted] cretinism, and **cachexia strumipirva** [redacted] two former developing in infancy or early childhood, [redacted] usually occurring in adult life. C. Cretinism may supervene after some infectious illness, such as **measles** or **whooping cough**. Sufferers from endemic cretinism are frequently dwarfs, from forty to sixty inches in height when fully grown, with short, broad bodies. The neck is short and thick, the abdomen large and pendulous, the forehead low, broad, and sloping, and the nose flat; the eyes are small and widely separated. The legs are short and sometimes crooked, with weak, ill developed muscles. Idiocy is an essential part of cretinism. Sporadic cretinism, unlike the endemic type, is not associated with goitre, but with a wasting or absence of the thyroid body; the outward signs are, however, very similar. It is characteristic of cretinism that the ossific centres are very late in appearing in the cartilaginous epiphyses, and the evidence of delayed endochondral ossification is well marked in the spine. The body is usually very diminutive, its true conjugate being two thirds and its breadth only half the normal dimensions. The chief features of the histology of cretinism are: a diminution in the amount of cartilage proliferation in the ossification zones, the formation of a continuous layer of bone separating the

2. Annotations

3. Entity records

Detected Entities

"goitre"

MANUAL

ENTITY ID
DOID:1459

NAMESPACE
DOID

STATUS FLAG
None (default)

Default: treat this entity as currently relevant.

POSITION IN TEXT
247 - 253

Remove Entity

4. Navigate and Save Buttons

Previous Next Save Current

Sample Index: 7 of 466 Go File: BMJ_1929_03_16_vol001_nr3558_art043_pmc2450088.txt ☐ Show Automatic Labels

1. First, hide the annotations that were made automatically.

are the organisms usually found in the throat. The blood shows a
sappear completely, but there is also a lymphocytic diminution, which
exhibit only slight changes. Blood cultures are positive in only 10 per cent.,
occus viridans, B. pyogenicus, B. cidii lactis, and B. coli. The disease is at
aged; it is apparently non contagious. The prognosis is bad, but not
serum and z ray applications to the long bones.

"agranulocytic angina"
MANUAL
ENTITY ID
DOID:2275
NAMESPACE
DOID
STATUS FLAG
Obsolete

1. Uncheck this box in the bottom-right corner of the page

001_nr3558_art043_pmc2450088.txt ▾

Sample 1 of 466

☐ Show Automatic Labels

2. To create a new manual annotation, select a text span with the mouse or touchpad. It can be single or multi-word, as below. A **Biomedical Entity** can be *disease, symptom, anatomy, phenotype, chemical entity, food material*, etc.. Do not include punctuation in the span.

File: BMJ_1929_03_16_vol001_nr3558_art043_pmc2450088.txt
Line: 11

W. C. HUEPER (Arch. Intern. Med., December, 1928, p. 893) recalls that Schultz, in 1922, described a type of necrotic angina accompanied by a marked absolute and especially **granulocytic leucopenia**, and regarded these symptoms as manifestations of a disease which he called "agranulocytosis." Since then about 125 cases have been recorded under this name or that of "**agranulocytic angina**" (Friedemann). Hueper now records observations on five cases seen between November, 1927, and April, 1928. The etiology is unknown. Most authorities regard it as an infectious disease represented by a **septicaemia** with an atypical reaction of the **haemopoietic system**, due either to bacteria with a special affinity and toxicity to the granulocytic system or to an atrophy and aplasia of this organ caused by septic infection. According to these investigations he had a case of a severe non septic septic or specific (septicocoele, B. pyogenicus, or fusosporilis). The disease starts after a period of prolonged ill health, or, more frequently, in previously healthy subjects, with high continued **fever**, **malaise**, **dyspnoea**, and dyspnoea; slight injuries is present in about 50 per cent. The patient rapidly gets worse, and death occurs after coma of two to seven days' duration; there may be rarely remissions of a few days to several weeks. The outcome is usually fatal. At the onset the **tonsils** are enlarged and reduced, and show yellowish white plugs, which merge to form dirty grey or yellowish coats; on removal of these an ulcerated surface appears. Sloughing of the tonsils rapidly ensues, and a similar necrotic process may be found on the **pharynx**, **uvula**, **palate**, **tongue**, pharynx, gums, **anus**, vulva, **vagina**, and pneumococci are the organisms usually found in the throat. The blood shows a **leucopenia** first and may disappear completely, but there is also a lymphocytic diminution, which is normal or exhibit only slight changes. Blood cultures are positive in only 10 per cent., Streptococcus viridans, B. pyogenicus, B. citri lactis, and B. coli. The disease is at the middle aged; it is apparently non contagious. The prognosis is bad, but not fatal. Anti streptococcus serum and x ray applications to the long bones.

In the text, select a span corresponding to a MEDICAL ENTITY

TYPES OF ENTITIES THAT YOU CAN ANNOTATE:

1. Anatomy
2. Cell type
3. Chemical / chemical exposure / environmental agent
4. Symptom
5. Phenotype
6. Genetic risk factors / inheritance / susceptibility
7. Onset (age, time of onset)
8. Taxonomy (of pathogens / organisms)
9. Transmission method (for infectious diseases)
10. Disease-driver / environmental-factor

To help simplify the annotation process, we would like to suggest the following:

If you are unsure which **fine-grained label** corresponds to an entity, but you can confidently identify a **broader category**, please feel free to use the broader label.

For example, instead of selecting a highly specific disease class, you may label the entity with one of the following higher-level terms:

"immune system disease" - DOID:2914

"disease by infectious agent" - DOID:0050117

"genetic disease" - DOID:630

"disease of metabolism" - DOID:0014667

"nutritional deficiency disease" - DOID:5113

"cognitive disorder" - DOID:1561

"lymphatic system disease"-DOID:75

etc..

If assigning even these broader categories is difficult, you may label the textual span simply as:

- *disease* — **DOID:4**

Similarly, higher-level categories are available for other entity types. These can be used whenever a more specific label cannot be confidently selected:

"symptom" - SYMP:0000462,

"anatomy" - UBERON:0001062,

"phenotype" - UPHENO:0001001,

"chemical entity" - CHEBI:24431,

"disease driver" - DISDRIV:0000000 (Environmental or genetic mechanisms driving the occurrence of complex diseases)

"evidence" - ECO:0000000 (A type of information that is used to support an assertion)

"food material" - FOODON:00002403

(this broad classes can also be used in the annotation in case a more fine-grained option cannot be found)

"transmission process" - TRANS_0000000

If you are a historian of medicine, you are encouraged to use broader labels for annotation.

3. When a span is selected, a **search window** will appear. It enables search of the entity in the Human Ontology of Disease.

NOTE: Search for a parent entity, if no exact entity can be identified: e.g., for “Streptococcus Viridans”, a parent entity can be “Streptococcus”

The screenshot shows a 'Add Entity Annotation' dialog box. At the top, it says 'Selected text: Streptococcus viridans'. Below this is a 'Search Ontology' input field containing 'Streptococcus viridans'. A blue arrow points to this field. To the right of the input field, a red box contains the text: 'Enter the entity's name—whether exact, partial, or approximate. You may also type a brief description or definition'. Below the search field, a list of suggestions is shown. The first suggestion is 'Streptococcus' with ID 'NCBITaxon:1301'. A blue arrow points to this suggestion. To the right of the list, a red box contains the text: 'The system will suggest the closest matching biomedical entity. Press to select the correct one.' Below the list, there is a text input field with the placeholder 'Or Enter' and a red box containing the text: 'Press to add annotation'. A blue arrow points to the 'Add Annotation' button at the bottom right.

Add Entity Annotation

Selected text:
Streptococcus viridans

Search Ontology
Streptococcus viridans

Fuzzy search enabled - results are ranked by

- ID NCBITaxon:1301
Streptococcus
- ID NCBITaxon:1314
Streptococcus pyogenes
- ID DOID:2898
obsolete commensal streptococcus
Synonyms: Streptococcal infection, Streptococcus infection
- ID DOID:0040084
Streptococcus pneumonia
A bacterial pneumonia has_material_basis_in Streptococcus pneumoniae.
- ID NCBITaxon:1313
Streptococcus pneumoniae

Or Enter
e.g., DOID:12345

Press to add annotation

Cancel Add Annotation

4. When you close the search window, you will see the annotation you just made on the right under “Detected entities.” Here, flag the entity as “historical”, “obsolete” or “None”. The differences between historical and obsolete are minor.

- *Obsolete terminology* refers to terms that are completely out of use today / adopted a different meaning.
- *Historical* means that it is still understood by modern medical professionals but used mainly in historical contexts.

The Human Disease Ontology also contains terms whose name starts with “obsolete” (type “obsolete” in the search field to see them). There, it serves for 1) the terms that do not represent a disease, these may include injuries, symptoms, phenotypes or terms that are no longer considered valid by the research community; 2) when two disease terms were merged into one. Example of an obsolete term due to not being a disease [Term] id: DOID:1000 name: Uterine Rupture synonym: "Rupture of uterus (disorder)"

Below you can find the instruction on where to flag the terms when you created an entity.

a) Find the entity record and the status flag:

Press on the annotated span
(in green) and the record of
the corresponding entity will
appear on the right

Find STATUS FLAG, click
on the drop-down menu

Detected Entities

"cretinous"

MANUAL

ENTITY ID
DOID:1059

NAMESPACE
DOID

STATUS FLAG
Obsolete

Obsolete or incorrect mapping. Ignore for downstream linking.

POSITION IN TEXT
210 - 219

Remove Entity

b) Select the appropriate flag in the drop-down menu

characterized by dwarf growth associated with
arents in districts where goitre and cretinism are
n is due to the insufficient supply of the secretion of the
tinct conditions—endemic cretinism, sporadic cretinism,
he latter, which follows the removal of the thyroid, usually

Select the appropriate flag
if the entity is obsolete

mea
n sh
ne e
t of cretinism. Sporadic cretinism, unlike the endemic
outward signs are, however, very similar. It is
inous epiphyses, and the evidence of delayed
, its true conjugate being two thirds and its breadth only

Detected Entities

"cretinous"

MANUAL

ENTITY ID

DOID:1059

NAMESPACE

DOID

None (default)

Historical

Obsolete

Obsolete or incorrect mapping. Ignore for downstream linking.

POSITION IN TEXT

210 – 219

Remove Entity

5. Continue assigning labels (and flagging) the entities, prioritizing the historical ones. Some items may already appear in green because another expert has annotated them, but please still add your own label—your vote helps us confirm whether several experts agree. If the text you want to select does not match exactly with what another expert highlighted, choose the wording or span that you believe is correct.

6. If you need to remove the labels that you have assigned incorrectly, **point** to the entity that you want to remove (in the text) with the mouse and go to the the entity record on the right:

with associated with
cre and cretinism are
apply of the secretion of the
etism, sporadic cretinism,
removal of the thyroid, usually
. Sufferers from endemic
eck is short and thick, the
separated. The legs are
nism, u
very s
dence of delayed

Detected Entities

"dwarf growth"

MANUAL

ENTITY ID
HP:0004322

NAMESPACE
HP

STATUS FLAG
Obsolete
Obsolete or incorrect mapping. Ignore for downstream link

POSITION IN TEXT
107 - 119

Remove Entity

Press this button to remove entity

7. When you are done annotating a text, press “Save Current” and then “Next” to move to the next text (bottom-left corner of the page).

endochondral ossification is well marked in the spine. The body is usually very diminutive, its true half the normal dimensions. The chief features of the histology of cretinism are: a diminution in ossification zones, the formation of a continuous layer of bone separating the

 **Instructions:**

- Select text with your mouse to add a new entity annotation
- Click on highlighted entities to view details
- Use the remove button to delete unwanted annotations
- Changes are saved automatically when you navigate

 Previous

Next 

 Save Current