

尿蛋白生物标志物数据库 标注应用案例

Medportal 是一个本体资源库，目前提供了 44 个本体，172 万多条术语。利用这些丰富的资源可以完成对项目相关术语的注释工作。本应用将以“尿蛋白生物标志物数据库”为例详细讲解使用 Medportal 提供的本体术语完成疾病名称注释工作的完整流程。

一、“尿蛋白生物标志物数据库”-UPBD 介绍

尿蛋白生物标志物数据库（Urinary Protein Biomarker Database, UPBD）提供了有关尿生物标志物或从已发表的文献汇编的候选生物标志物的信息。该数据库包含了对患者或实验动物尿液标本的蛋白质组学和非蛋白质组学的研究。一旦一种蛋白质或肽被报道在疾病组和对照组之间或在不同疾病阶段之间表达不同，它就被视为候选生物标志物。通过该数据库的应用，您可以更全面地了解泌尿系生物标志物及其相关疾病的发病机制，为进一步的深入研究提供有益参考。本文档重点介绍如何使用 Medportal 中的本体术语来对 UPBD 数据库中的疾病名称进行注释。

示例如下：

1) 选择 UPBD 中的疾病名称

Disease	Treatment	Biomarker usage	Protein name	Organism	Specimen	In a Panel	
1 Bacterial sepsis of new		Diagnosis	C-C motif chemokine 2	Human	Urine	False	More
2 Bacterial sepsis of new		Diagnosis	C-X-C motif chemokine	Human	Urine	False	More
3 Bacterial sepsis of new		Diagnosis	Interleukin-8	Human	Urine	False	More
4 Schistosomiasis		Diagnosis	C-C motif chemokine 2	Human	Urine	False	More
5 Sepsis due to Enterococ		Diagnosis	Neutrophil gelatinase-	Human	Urine	False	More
6 Visceral leishmaniasis		Diagnosis	Kininogen-1	Human	Urine	False	More

2) 点击链接，查看 Medportal 中相应本体提供的术语名称、定义等注释信息。

Details	Visualization	Notes (0)	Class Mappings (3)
Preferred Name	schistosomiasis		
Definitions	A parasitic helminthiasis infectious disease that involves infection of the intestine, urinary tract, skin, liver and spleen caused by multiple species of the trematode fluke of the genus Schistosoma. The symptoms include fever, chills, nausea, abdominal pain, diarrhea, malaise, myalgia, liver and spleen enlargement, rash and hematuria.		
ID	http://purl.obolibrary.org/obo/DOID_1395		
database_cross_reference	SNOMEDCT_US_2016_03_01:187116001 ICD9CM:120.8 SNOMEDCT_US_2016_03_01:187121003 UMLS_CUI:C0029827		
definition	A parasitic helminthiasis infectious disease that involves infection of the intestine, urinary tract, skin, liver and spleen caused by multiple species of the trematode fluke of the genus Schistosoma. The symptoms include fever, chills, nausea, abdominal pain, diarrhea, malaise, myalgia, liver and spleen enlargement, rash and hematuria.		
has_obo_namespace	disease_ontology		
id	DOID:1395		
label	schistosomiasis		
notation	DOID:1395		
prefLabel	schistosomiasis		
treeView	parasitic helminthiasis infectious disease		
subClassOf	parasitic helminthiasis infectious disease		

二、使用 Medportal 完成对 UPBD 的注释

2.1 需求和分析

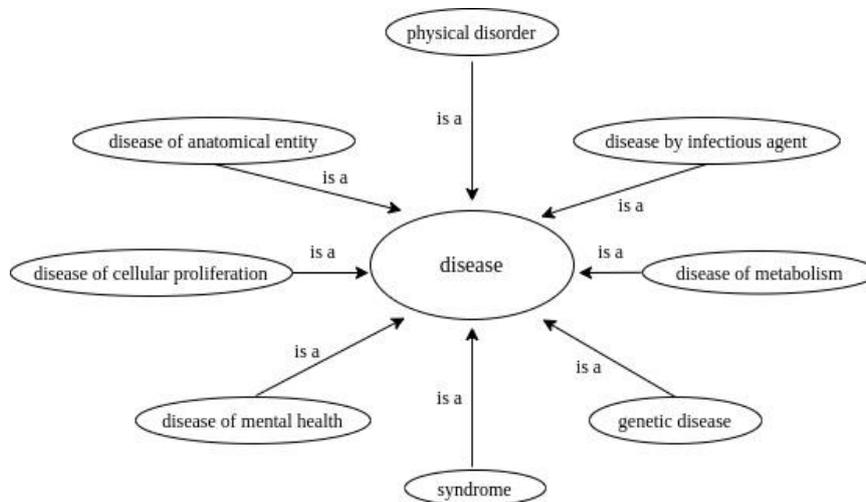
- ☑支持尿蛋白生物标志物研究的语义注释。
- ☑目前还没有单个本体可以满足需求，需要多本体联合才能支持完整的语义结构。

2.2 设计原则

- ☑遵循 OBO Foundry 原则。
- ☑尽可能的重用 Medportal 本体中的术语。
- ☑使用疾病本体(Disease Ontology, DO) (<https://disease-ontology.org/>) 作为主术语集并集成其它本体中的术语。DO 本体的 ID 为 DOID。

2.3 为什么选择 DO 作为主术语集

- ☑它覆盖了疾病的主要方面。
- ☑包含了对其他临床术语的交叉引用(如: MESH、OMIM、ICD9CM、ICD10CM、NCI Thesaurus、SNOMED CT、Orphanet-Rare Diseases 等)。



DO 疾病分类结构图

2.4 实现过程

2.4.1 整理尿蛋白生物标志物关联的疾病名称

作为示例，我们只选取了尿蛋白生物标志物关联的 5 种参考疾病，如表 1 所示：

表 1 尿蛋白生物标志物关联的参考疾病名称

原始疾病名称
Schistosomiasis
Bacterial sepsis of newborn
Kala-Azar
Sepsis due to Enterococcus
Abnormality of the nephron

2.4.2 查询 Medportal, 在本体中寻找疾病对应的术语

在 Medportal 中, 有三种方式可以查询本体术语, 分别为: Search 工具、Annotator 工具和 API 编程访问接口。用户可根据需要选择其中一种或多种完成查询工作。

2.4.2.1 使用 Medportal Search 工具进行查询

- 查询 Schistosomiasis

MedPortal Ontologies Search Annotator Recommender Mappings Admin admin Support

Class Search

Schistosomiasis

Enter a class, e.g. Melanoma [help](#)

[Show advanced options](#)

[Search](#)

Matches in 7 ontologies

[schistosomiasis - Human Disease Ontology \(DOID\)](#)
http://purl.obolibrary.org/obo/DOID_1395
 A parasitic helminthiasis infectious disease that involves infection of the intestine, urinary tract, skin, liver and spleen caused by multiple species of the trematode fluke of the genus Schistosoma. The ...
[details](#) - [visualize](#) - 3 more from this ontology

[Schistosomiasis - Medical Subject Headings \(MESH\)](#)
<http://purl.bioontology.org/ontology/MESH/DO12252>
 Infection with flukes (trematodes) of the genus SCHISTOSOMA. Three species produce the most frequent clinical diseases: SCHISTOSOMA HAEMATOBIIUM (endemic in Africa and the Middle East), SCHISTOSOMA MANSONI (in ...
[details](#) - [visualize](#) - 4 more from this ontology

从查询结果可以看出, 该疾病存在于 7 个本体中。由于我们首选 DO 本体术语集, 所以此处选择 DO 中的术语作为注释信息。

- 查询 Bacterial sepsis of newborn

MedPortal Ontologies Search Annotator Recommender Mappings [Login](#) Support

Class Search

Bacterial sepsis of newborn

Enter a class, e.g. Melanoma [help](#)

[Show advanced options](#)

[Search](#)

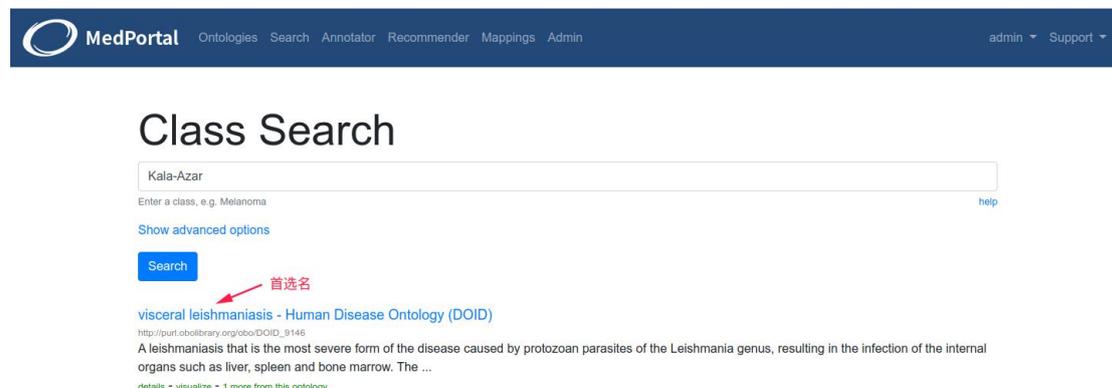
Matches in 2 ontologies

[Bacterial sepsis of newborn - International Classification of Diseases, Version 10 - Clinical Modification \(ICD10CM\)](#)
<http://purl.bioontology.org/ontology/ICD10CMP36>
[details](#) - [visualize](#)

[Bacterial sepsis of newborn - International Classification of Diseases, Version 10 \(ICD10\)](#)
<http://purl.bioontology.org/ontology/ICD10/P36>
[details](#) - [visualize](#)

从查询结果可以看出，该疾病存在于 ICD10CM 和 ICD10 两个本体中。此时我们可以根据需要自行选择一个本体中的术语作为注释信息（本示例中我们选择了 ICD10CM 中的术语作为注释信息）。

● 查询 Kala-Azar



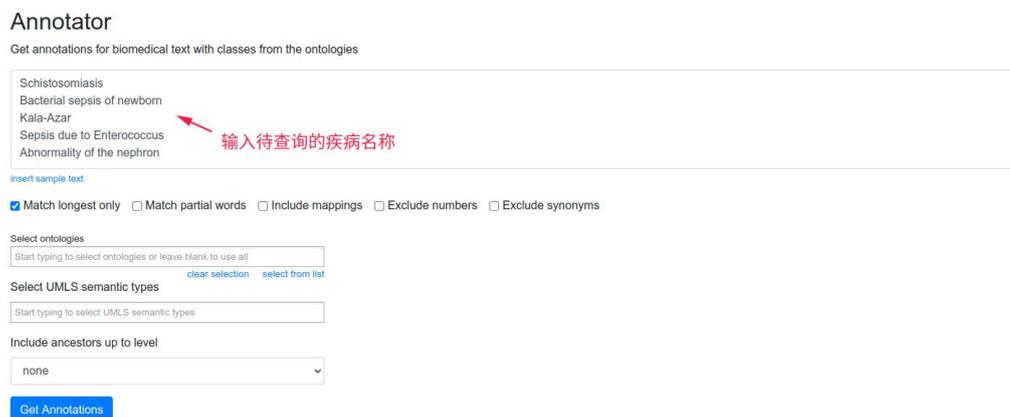
查询结果显示的术语名称是 visceral leishmaniasis（而不是 Kala-Azar）。这是因为 Kala-Azar 是 visceral leishmaniasis 的同义词，而 visceral leishmaniasis 是首选名，所以优先显示。该术语对应的完整注释如下图所示：

Details	Visualization	Notes (0)	Class Mappings (3)
Preferred Name	visceral leishmaniasis		
Synonyms	Kala-Azar Infection by visceral leishmaniasis		
Definitions	Xref MGI. A leishmaniasis that is the most severe form of the disease caused by protozoan parasites of the Leishmania genus, resulting in the infection of the internal organs such as liver, spleen and bone marrow. The symptoms include fever, weight loss, and an enlarged spleen and liver.		
ID	http://purl.obolibrary.org/obo/DOID_9146		

其他两个疾病名称，读者自行查询，这里不再赘述。

2.4.2.2 使用 Medportal Annotator 工具进行查询

Medportal Annotator 提供了使用本体术语对文本进行注释的 web 服务。为了获取疾病名称对应的本体术语，需将每个疾病名称作为单独一行，依次写入到 Annotator 的查询文本框中。点击“Get Annotations”按钮后，每一个被查询的疾病名称对应的类名(术语)、本体名都将在查询结果中显示出来。



查询结果

Annotations

total results 11 (direct 11 / ancestor 0 / mapping 0)

CLASS	filter	ONTOLOGY	filter	TYPE	filter	ONTEXT	MATCHED CLASS	filter	MATCHED ONTOLOGY	filter
schistosomiasis		Human Disease Ontology		direct	Schistosomiasis	Bacterial sepsis of newborn Kala-Azar Sepsis ...	schistosomiasis		Human Disease Ontology	
Schistosomiasis		National Cancer Institute Thesaurus		direct	Schistosomiasis	Bacterial sepsis of newborn Kala-Azar Sepsis ...	Schistosomiasis		National Cancer Institute Thesaurus	
Schistosomiasis		Medical Subject Headings		direct	Schistosomiasis	Bacterial sepsis of newborn Kala-Azar Sepsis ...	Schistosomiasis		Medical Subject Headings	
Bacterial sepsis of newborn		International Classification of Diseases, Version 10 - Clinical Modification		direct	Schistosomiasis	Bacterial sepsis of newborn Kala-Azar Sepsis due to Enterococcus Abnormality ...	Bacterial sepsis of newborn		International Classification of Diseases, Version 10 - Clinical Modification	
Bacterial sepsis of newborn		International Classification of Diseases, Version 10		direct	Schistosomiasis	Bacterial sepsis of newborn Kala-Azar Sepsis due to Enterococcus Abnormality ...	Bacterial sepsis of newborn		International Classification of Diseases, Version 10	
visceral leishmaniasis		Human Disease Ontology		direct	Schistosomiasis ... sepsis of newborn Kala-Azar Sepsis due to Enterococcus Abnormality ...		visceral leishmaniasis		Human Disease Ontology	
Kala-azar		MedlinePlus Health Topics		direct	Schistosomiasis ... sepsis of newborn Kala-Azar Sepsis due to Enterococcus Abnormality ...		Kala-azar		MedlinePlus Health Topics	
Kala-Azar		National Cancer Institute Thesaurus		direct	Schistosomiasis ... sepsis of newborn Kala-Azar Sepsis due to Enterococcus Abnormality ...		Kala-Azar		National Cancer Institute Thesaurus	

2.4.2.3 使用 Medportal API 进行查询

为了方便开发人员，Medportal 提供了应用程序接口(API)，开发人员可以通过编程方式查询需要的结果。API 查询过程如下：

- 1) 在 Medportal 网站注册用户。
- 2) 登录后获取到该用户的 API。

API Key Your API Key can be used to access the NCBO API (REST) Services

37bcf440-...-700e52e

- 3) 该 API 中关于术语查询的说明如下

Term Search

An endpoint for searching terms and their properties is available

- Perform search
 - **GET** **POST** /search?q={search query}
 - example: /search?q=melanoma
- Parameters
 - ontologies={ontology_id1,ontology_id2,ontology_id3}
 - require_exact_match={true|false} // default = false
 - suggest={true|false} // default = false. Will perform a search specifically geared towards type-ahead suggestions.
 - also_search_views={true|false} // Include ontology views in the search. default = false
 - require_definitions={true|false} // default = false
 - also_search_properties={true|false} // default = false
 - also_search_obsolete={true|false} // default = false (exclude obsolete terms)
 - cui={C0018787,C0225807,C0018787} // Allows restricting query by CUIs. default = null (no restriction)
 - semantic_types={T023,T185,T061} // Allows restricting query by Semantic Types (TUI). default = null (no restriction)
 - include={prefLabel, synonym, definition, notation, cui, semanticType} // default = (see Common Parameters section)
 - page={integer representing the page number} // default = 1
 - pagesize={integer representing the size of the returned page} // default = 50

- 4) 举例：查询疾病名称“Bacterial sepsis of newborn”

在编程中，开发人员需要指定如下的访问链接：

[http://data.medportal.bmicc.cn/search?q=Bacterial sepsis of newborn&apikey=yourapikey&require_exact_match=true](http://data.medportal.bmicc.cn/search?q=Bacterial+sepsis+of+newborn&apikey=yourapikey&require_exact_match=true)

从该链接中获取的返回结果如下：

```
{
  page: 1,
  pageCount: 1,
  totalCount: 2,
  prevPage: null,
  nextPage: null,
  links: {
    nextPage: null,
    prevPage: null
  },
  collection: [
    - {
      prefLabel: "Bacterial sepsis of newborn",
      synonym: [
        "congenital sepsis"
      ],
      cui: [
        "C1260896",
        "C3665339"
      ],
      semanticType: [
        "T047"
      ],
      obsolete: false,
      matchType: "prefLabel",
      ontologyType: "ONTOLOGY",
      provisional: false,
      @id: http://purl.bioontology.org/ontology/ICD10CM/P36,
      @type: http://www.w3.org/2002/07/owl#Class,
      links: {
        self: http://data.medportal.bmicc.cn/ontologies/ICD10CM/classes/http%3A%2Fpurl.bioontology.org%2Fontology%2FICD10CM%2FP36,
        ontology: http://data.medportal.bmicc.cn/ontologies/ICD10CM,
        children: http://data.medportal.bmicc.cn/ontologies/ICD10CM/classes/http%3A%2Fpurl.bioontology.org%2Fontology%2FICD10CM%2FP36/children,
        parents: http://data.medportal.bmicc.cn/ontologies/ICD10CM/classes/http%3A%2Fpurl.bioontology.org%2Fontology%2FICD10CM%2FP36/parents,
        descendants: http://data.medportal.bmicc.cn/ontologies/ICD10CM/classes/http%3A%2Fpurl.bioontology.org%2Fontology%2FICD10CM%2FP36/descendants,
        ancestors: http://data.medportal.bmicc.cn/ontologies/ICD10CM/classes/http%3A%2Fpurl.bioontology.org%2Fontology%2FICD10CM%2FP36/ancestors,
        instances: http://data.medportal.bmicc.cn/ontologies/ICD10CM/classes/http%3A%2Fpurl.bioontology.org%2Fontology%2FICD10CM%2FP36/instances,
      }
    }
  ]
}
```

上例中，q 参数后面是查询内容，require_exact_match=true 表明需要完全匹配。由于示例中没有指明从哪个本体中查询，所以它将默认查询所有本体。如果要明确指出查询的本体，例如：从 DOID 和 ICD10CM 中查询，则可以使用 ontologies=DOID,ICD10CM。

查询结果 page 告诉我们一共有多少页，pageCount 说明每页有多少条记录，totalCount 指出一共有多少条记录。

2.4.3 整理查询结果

在使用前面的各种查询方法，检索到疾病名称对应的本体术语后，可将其整理成表 2 所示的样子：

表 2 原始疾病名称和本体术语的映射关系

原始疾病名称	对应术语	对应本体	术语 ID
Schistosomiasis	schistosomiasis	DOID	http://purl.obolibrary.org/obo/DOID_1395
Bacterial sepsis of newborn	Bacterial sepsis of newborn	ICD10CM	http://purl.bioontology.org/ontology/ICD10CM/P36
Kala-Azar	visceral leishmaniasis	DOID	http://purl.obolibrary.org/obo/DOID_9146
Sepsis due to Enterococcus	Sepsis due to Enterococcus	ICD10CM	http://purl.bioontology.org/ontology/ICD10CM/A41.81
Abnormality of the nephron	Abnormality of the nephron	HP	http://purl.obolibrary.org/obo/HP_0012575

2.4.4 使用查询结果完成对疾病名称的注释

前面整理的查询结果，实际上在疾病名称和本体术语之间建立了映射关系。通过术语的首选名称、可选名称、同义词、定义等相关信息可以实现对疾病的详细注释。注释链接的格式为：

```
http://medportal.bmicc.cn/ontologies/ontologyID/?p=classes&conceptid=termID
```

- 说明：
- 1) ontologyID 为本体的 ID，例如 DO 本体的 ID 为 DOID。
 - 2) p=classes 表明展示的是类(术语)的详细信息。
 - 3) termID 为类的 ID，即：类的标识符。

举例：点击疾病名称 Schistosomiasis 的超链接可以获得注释信息

Disease	Treatment	Biomarker usage	Protein name	Organism	Specimen	In a Panel
Schistosomiasis						
1 Schistosomiasis		Diagnosis	C-C motif chemokine ; Human		Urine	False

链接为：

http://medportal.bmicc.cn/ontologies/DOID/?p=classes&conceptid=http://purl.obolibrary.org/obo/DOID_1395

获取到的注释内容为：

Details	Visualization	Notes (0)	Class Mappings (3)
Preferred Name	schistosomiasis		
Definitions	A parasitic helminthiasis infectious disease that involves infection of the intestine, urinary tract, skin, liver and spleen caused by multiple species of the trematode fluke of the genus Schistosoma. The symptoms include fever, chills, nausea, abdominal pain, diarrhea, malaise, myalgia, liver and spleen enlargement, rash and hematuria.		
ID	http://purl.obolibrary.org/obo/DOID_1395		
database_cross_reference	SNOMEDCT_US_2016_03_01:187116001 ICD9CM:120.8 SNOMEDCT_US_2016_03_01:187121003 UMLS_CUI:C0029827		
definition	A parasitic helminthiasis infectious disease that involves infection of the intestine, urinary tract, skin, liver and spleen caused by multiple species of the trematode fluke of the genus Schistosoma. The symptoms include fever, chills, nausea, abdominal pain, diarrhea, malaise, myalgia, liver and spleen enlargement, rash and hematuria.		
has_obo_namespace	disease_ontology		
id	DOID:1395		
label	schistosomiasis		
notation	DOID:1395		
prefLabel	schistosomiasis		
treeView	parasitic helminthiasis infectious disease		
subClassOf	parasitic helminthiasis infectious disease		

至此，利用 Medportal 中本体提供的术语来完成对 UPBD 数据库中疾病名称的注释就介绍完了。