

科研图谱

图谱介绍

科研图谱主要是从病毒分类、病毒基本信息、抗病毒药物、病毒亲缘关系图谱等角度进行科研分析，具体包括四个子图谱：

病毒分类图谱

病毒分类图谱是根据最新ncbi网站内容更新了病毒族系相关的科研基础数据，其中以新冠病毒科为主，主要是对病毒从门、纲、目、科、属等进行分类分析。

新冠基本信息图谱

从新冠病毒的基因蛋白，宿主等相关信息出发构图

抗病毒药物图谱

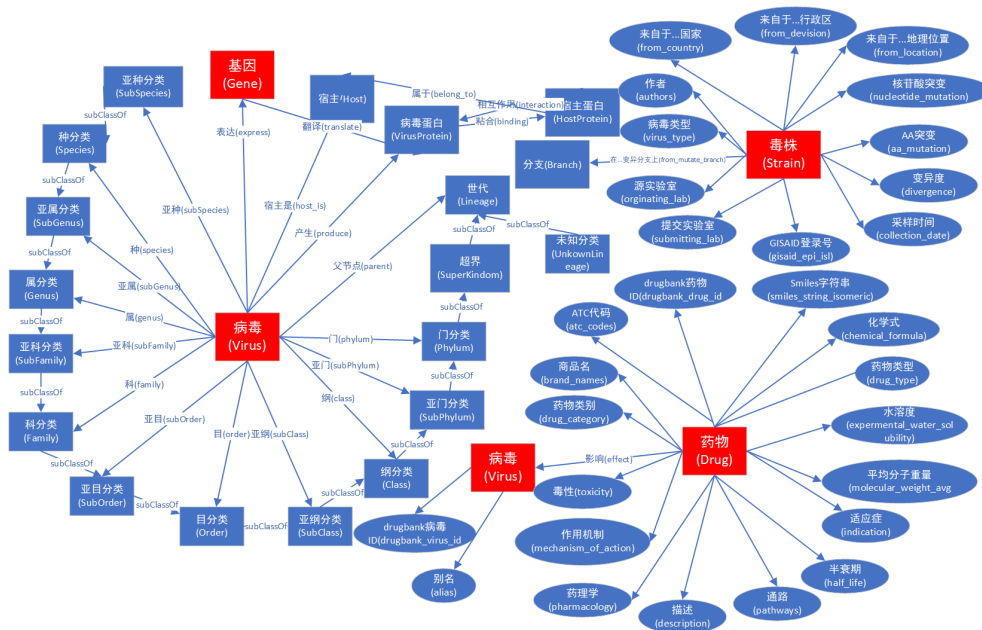
依据抗病毒药物、病毒、病毒相关蛋白和宿主及宿主蛋白间的关系构建图谱

病毒亲缘关系图谱

整理短时间内COVID-19病毒株的变异方向与地理位置分布数据

图谱 Schema

如图所示



其中，矩形框为实体，椭圆形为属性

图谱数据

目前已经分别构建了如下知识图谱

- 病毒分类图谱（中文，英文）

包括中文知识图谱和英文知识图谱，共两个知识图谱(名称为:病毒分类图谱-中文、病毒分类图谱-英文)

三元组数量：1934954

实体数量：205633

谓词数量：16（括号里面的为英文谓词）

- 门(phylum)
- 亚门(subPhylum)
- 属(genus)
- 亚属(subGenus)
- 科(family)
- 亚科(subfamily)
- 纲(class)
- 目(order)
- 亚目(subOrder)
- 父节点(parent)
- 种(species)
- domain(domain)
- identifier(identifier)
- range(range)
- <http://www.w3.org/2000/01/rdf-schema#label>(<http://www.w3.org/2000/01/rdf-schema#label>)
- <http://www.w3.org/1999/02/22-rdf-syntax-ns#type>(<http://www.w3.org/1999/02/22-rdf-syntax-ns#type>)

- 新冠基本信息图谱

包括中文知识图谱和英文知识图谱，共两个知识图谱(名称为:新冠基本信息图谱-中文、新冠基本信息图谱-英文)

三元组数量: 2879

实体数量: 553

谓词数量: 10 (括号里面的为英文谓词)

- 翻译(translate)
- 宿主是(host_is)
- 产生(produce)
- 表达(express)
- domain(domain)
- identifier(identifier)
- range(range)
- url(url)
- <http://www.w3.org/2000/01/rdf-schema#label>(<http://www.w3.org/2000/01/rdf-schema#label>)
- <http://www.w3.org/1999/02/22-rdf-syntax-ns#type>(<http://www.w3.org/1999/02/22-rdf-syntax-ns#type>)

• 抗病毒药物图谱

包括中文知识图谱和英文知识图谱，共两个知识图谱(名称为:病毒药物图谱-中文、病毒药物图谱-英文)

三元组数量: 61796

实体数量: 7999

谓词数量: 29 (括号里面的为英文谓词)

- drug_category(drug_category)
- brand_names(brand_names)
- drugbank_drug_id(drugbank_drug_id)
- drug_type(drug_type)
- act_codes(act_codes)
- description(description)
- pathways(pathways)
- half_life(half_life)
- pharmacology(pharmacology)
- smiles_string_isomeric(smiles_string_isomeric)
- uniprotkb_entry_name(uniprotkb_entry_name)
- chemical_formula(chemical_formula)
- alias(alias)
- mechanism_of_action(mechanism_of_action)
- molecular_weight_avg(molecular_weight_avg)
- experimental_water_solubility(experimental_water_solubility)
- toxicity(toxicity)
- domain(domain)
- identifier(identifier)
- range(range)
- <http://www.w3.org/2000/01/rdf-schema#label>(<http://www.w3.org/2000/01/rdf-schema#label>)
- <http://www.w3.org/1999/02/22-rdf-syntax-ns#type>(<http://www.w3.org/1999/02/22-rdf-syntax-ns#type>)
- 属于(belong_to)
- 影响(effect)
- 相互作用(interaction)
- 产生(produce)
- 粘合(binding)

• 病毒亲缘关系图谱

包括中文知识图谱和英文知识图谱，共两个知识图谱(名称为:病毒亲缘关系图谱-中文、病毒亲缘关系图谱-英文)

三元组数量: 2237

实体数量: 464

谓词数量: 17 (括号里面的为英文谓词)

- 来自于...国家(from_country)
- 来自于...行政区(from_division)
- 来自于...地理位置(from_location)
- 宿主是(host_is)
- 病毒类型(virus_type)
- 源实验室(originating_lab)
- 提交实验室(submitting_lab)
- 作者们(authors)
- 在...变异分支上(from_mutate_branch)
- 变异度(divergence)
- AA突变(aa_mutation)
- GISAID登录号(gisaid_epi_isl)
- 核苷酸突变(nucleotide_mutation)
- domain(domain)
- range(range)
- <http://www.w3.org/2000/01/rdf-schema#label>(<http://www.w3.org/2000/01/rdf-schema#label>)
- <http://www.w3.org/1999/02/22-rdf-syntax-ns#type>(<http://www.w3.org/1999/02/22-rdf-syntax-ns#type>)

• 新冠科研图谱

同时我们还将四个子知识图谱进行了合并，合并成了新冠科研图谱(共两个:新冠科研图谱-中文、新冠科研图谱-英文)

三元组数量: 2000928

实体数量: 214400

谓词数量: 57 (谓词为四个知识图谱谓词的并集)

图谱查询示例

1. 查询指定病毒（Bovine adenovirus 4）的分类信息（科、属、种信息）

中文知识图谱SPARQL语句（数据库为：病毒分类图谱-中文）

```
select ?x ?c11 ?c21 ?c31 where
{
  ?x <http://www.w3.org/2000/01/rdf-schema#label> "Bovine adenovirus 4"@en.
  ?x <科> ?c1.
  ?c1 <http://www.w3.org/2000/01/rdf-schema#label> ?c11.
  ?x <种> ?c2.
  ?c2 <http://www.w3.org/2000/01/rdf-schema#label> ?c21.
  ?x <属> ?c3.
  ?c3 <http://www.w3.org/2000/01/rdf-schema#label> ?c31.
}
```

英文知识图谱SPARQL语句（数据库为：病毒分类图谱-英文）

```
select ?x ?c11 ?c21 ?c31 where
{
  ?x <http://www.w3.org/2000/01/rdf-schema#label> "Bovine adenovirus 4"@en.
  ?x <family> ?c1.
  ?c1 <http://www.w3.org/2000/01/rdf-schema#label> ?c11.
  ?x <species> ?c2.
  ?c2 <http://www.w3.org/2000/01/rdf-schema#label> ?c21.
  ?x <genus> ?c3.
  ?c3 <http://www.w3.org/2000/01/rdf-schema#label> ?c31.
}
```

2. 查询指定药品（Elvitegravir）的相关信息

中文知识图谱SPARQL语句（数据库为：病毒药物图谱-中文）

```
select ?x ?k ?p ?z ?y where
{
  ?x <http://www.w3.org/2000/01/rdf-schema#label> "Elvitegravir"@en.
  ?x <drug_category> ?k.
  ?x <alias> ?p.
  ?x <影响> ?z.
  ?x <atc_codes> ?y.
}
```

英文知识图谱SPARQL语句（数据库为：病毒药物图谱-英文）

```
select ?x ?k ?p ?z ?y where
{
  ?x <http://www.w3.org/2000/01/rdf-schema#label> "Elvitegravir"@en.
  ?x <drug_category> ?k.
  ?x <alias> ?p.
  ?x <effect> ?z.
  ?x <atc_codes> ?y.
}
```

3. 查询指定药物（Elvitegravir）影响病毒信息

中文知识图谱SPARQL语句（数据库为：新冠科研图谱-中文）

```
select ?x ?k ?p ?z ?y ?p1 ?o where
{
  ?x <http://www.w3.org/2000/01/rdf-schema#label> "Elvitegravir"@en.
  ?x <drug_category> ?k.
  ?x <alias> ?p.
  ?x <影响> ?z.
  ?x <atc_codes> ?y.
  ?z ?p1 ?o.
}
```

英文知识图谱SPARQL语句（数据库为：新冠科研图谱-英文）

```
select ?x ?k ?p ?z ?y ?p1 ?o where
{
  ?x <http://www.w3.org/2000/01/rdf-schema#label> "Elvitegravir"@en.
  ?x <drug_category> ?k.
  ?x <alias> ?p.
  ?x <effect> ?z.
  ?x <atc_codes> ?y.
  ?z ?p1 ?o.
}
```

4. 查询病毒（SARS-CoV-2）的基因信息

中文知识图谱SPARQL语句（数据库为：新冠科研图谱-中文）

```
select ?x ?y ?y1 where
{
  ?x <http://www.w3.org/2000/01/rdf-schema#label> "SARS-CoV-2"@en.
  ?x <表达> ?y.
  ?y <http://www.w3.org/1999/02/22-rdf-syntax-ns#type> ?z.
  ?z <http://www.w3.org/2000/01/rdf-schema#label> "基因"@zh.
  ?y <http://www.w3.org/2000/01/rdf-schema#label> ?y1.
}
```

英文知识图谱SPARQL语句（数据库为：新冠科研图谱-英文）

```
select ?x ?y ?y1 where
{
  ?x <http://www.w3.org/2000/01/rdf-schema#label> "SARS-CoV-2"@en.
```

```
?x <express> ?y.  
?y <http://www.w3.org/1999/02/22-rdf-syntax-ns#type> ?z.  
?z <http://www.w3.org/2000/01/rdf-schema#label> "基因"@zh.  
?y <http://www.w3.org/2000/01/rdf-schema#label> ?y1.  
}
```

5. 查询可以影响（或抑制）病毒的药物有那些？

```
select ?x ?y ?x1 where  
{  
  ?x <影响> ?y.  
  ?y <http://www.w3.org/2000/01/rdf-schema#label> "Human immunodeficiency virus 1"@en.  
  ?x <http://www.w3.org/2000/01/rdf-schema#label> ?x1.  
}
```

图谱API地址和参数示例

系统提供了基于HTTP请求的API访问接口，可以实现基于SPARQL的远程数据查询

```
API地址: http://cloud.gstore.cn/api  
  
AccessKeyID:08745175bc8c43de8e981706a4c1df6a  
  
Secret:D8B3D8179CD1F857F61BD406F3690CBA
```

具体参考API帮助文档<http://cloud.gstore.cn/gStoreAPI.html>