**一、目前已有的数据源**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 数据源 | 文件数 | 大小 | 格式 | 描述 |
| TCMGeneDIT | 7 | 15.7MB | Rdf | tcm-disease  tcm-effect  tcm-gene  tcm-gene-disease  tcm-ingredient  gene-name-mapping |
| Gene2go | 1 | 131MB | Rdf | 基因entrez-id与go-id的对应 |
| FunDO | 1 | 1.28MB | Rdf | 疾病与基因联系 |
| Gene Ontology | 1 | 129MB | Rdf | 基因本体 |
| Disease Ontology | 1 | 43.1MB | Rdf | 疾病本体 |
| Disease name-id | 1 | 191kb | Rdf | 疾病名与id对应 |
| Uniprot protein-gene | 1 | 11.3GB | rdf | 蛋白质-基因联系 |
| Diseasesome |  |  |  |  |
| Drugbank |  |  |  |  |
| dailymed |  |  |  |  |

**二、数据源之间的联系**

Uniprot protein-gene

TCMGeneDIT

Entrez-id

Entrez-id

Gene2Go

GO-id

GO-id

Disease name

Entrez-id

Gene Ontology

Disease name-id

Disease

name

FunDO

Entrez-id

Gene symbol-id

Disease id

Disease Ontology

Gene symbol

Diseasesome

Drug id

Drug id

Drugbank

Dailymed

**三、各个数据源内部结构**

1、TcmGeneDIT

j.0: treatment

j.0: effect

j.0: association

j.0:contextAssociation

j.0:contextAssociation

j.0:contextAssociation

j.0:ingredient

owl:sameAs

注：

1）图中实体的URI模式

Medicine：http://purl.org/net/tcm/tcm.lifescience.ntu.edu.tw/id/medicine/*name*

Gene：http://purl.org/net/tcm/tcm.lifescience.ntu.edu.tw/id/gene/*id*

Disease：http://purl.org/net/tcm/tcm.lifescience.ntu.edu.tw/id/disease/*name*

Ingredient：http://purl.org/net/tcm/tcm.lifescience.ntu.edu.tw/id/ingredient/*name*

Effect：http://purl.org/net/tcm/tcm.lifescience.ntu.edu.tw/id/effect/*name*

Association：http://purl.org/net/tcm/tcm.lifescience.ntu.edu.tw/id/association/*id*

2）这些实体每一个都包含属性rdf:type以及rdfs:label。

3）该数据库中还包含实体http://purl.org/net/tcm/tcm.lifescience.ntu.edu.tw/id/statistics/*id*，每一个这样的实体包含了一个属性j.0:medicine\_effect\_association\_tvalue，它指明了上图中每一个关联关系的关联度。这个关联值目前暂时不用。

4）xmlns:j.0="http://purl.org/net/tcm/tcm.lifescience.ntu.edu.tw/

2、gene2go

http://purl.org/net/tcm/tcm.lifescience.ntu.edu.tw/association

3、FunDO

http://purl.org/net/tcm/tcm.lifescience.ntu.edu.tw/association

4、Gene Ontology

Rdfs:label

Rdfs:subClassof

oboInOwl:

hasOBONamespace

oboInOwl:hasExactSynonym

oboInOwl:id

oboInOwl:

hasDBXref

注：

1）命名空间xmlns:go="http://www.geneontology.org/dtds/go.dtd#"

2）term的uri模式为http://purl.obolibrary.org/obo/GO\_*id*

部分属性未列出。

5、Disease Ontology

Rdfs:label

oboInOWL:hasExactSynonym

oboInOWL:xref

Rdfs:subClassOf

注：term：http://purl.obolibrary.org/obo/DOID\_*id*

6、Disease name id mapping

owl:sameAs

注：

Name uri有两种（分别来自TcmGeneDIT和FunDO）

http://purl.org/net/tcm/tcm.lifescience.ntu.edu.tw/id/disease/*name*

http://biotcmcloud.org/disease/*name*

7、Uniprot protein-gene

注：j.0=” http://purl.uniprot.org/core/”

j.0: classifiedWith

j.0: classifiedWith

j.0: mnemonic

8、diseasesome

http://www4.wiwiss.fu-berlin.de/diseasome/resource/diseasome/name

http://www4.wiwiss.fu-berlin.de/diseasome/resource/diseasome/associatedGene

http://www4.wiwiss.fu-berlin.de/diseasome/resource/diseasome/possibleDrug

http://www4.wiwiss.fu-berlin.de/diseasome/resource/diseasome/possibleDrug

9、drugbank

http://www.w3.org/2000/01/rdf-schema#label

http://www4.wiwiss.fu-berlin.de/drugbank/resource/drugbank/text

http://www4.wiwiss.fu-berlin.de/drugbank/resource/drugbank/interactionDrug2

http://www4.wiwiss.fu-berlin.de/drugbank/resource/drugbank/interactionDrug1

http://www4.wiwiss.fu-berlin.de/drugbank/resource/drugbank/target

http://www.w3.org/2000/01/rdf-schema#label

http://www4.wiwiss.fu-berlin.de/drugbank/resource/drugbank/hgncId

http://www4.wiwiss.fu-berlin.de/drugbank/resource/drugbank/name

http://www4.wiwiss.fu-berlin.de/drugbank/resource/drugbank/geneSequence

http://www4.wiwiss.fu-berlin.de/drugbank/resource/drugbank/proteinSequence

http://www4.wiwiss.fu-berlin.de/drugbank/resource/drugbank/swissprotId

10、dailymed

http://www4.wiwiss.fu-berlin.de/dailymed/resource/dailymed/description

http://www4.wiwiss.fu-berlin.de/dailymed/resource/dailymed/inactiveIngredient

http://www.w3.org/2000/01/rdf-schema#label