知识工程项目1：COVID-19知识图谱系统开发

本项目的总体要求为使用指定的COVID-19知识图谱进行系统开发，实现知识图谱数据存储、查询、挖掘与可视化展示。

1. 知识图谱数据集介绍

COVID-19知识图谱：COVID-19 Knowledge Graph

网站：<https://research.bioinformatics.udel.edu/covid19kg>

论文：<https://academic.oup.com/bioinformatics/article/37/23/4597/6382302>

SPARQL查询界面：<https://research.bioinformatics.udel.edu/covid19kg/yasgui>

RESTful API接口：<https://research.bioinformatics.udel.edu/covid19kg/api>

RDF文件下载：<https://research.bioinformatics.udel.edu/covid19kg/rdfdumps>

1. 系统开发要求
2. 开发具有后台知识图谱数据库和前台用户界面的COVID-19知识图谱管理信息系统。
3. 后台开发要求：从COVID-19知识图谱中选取RDF数据集下载，导入知识图谱数据库进行存储管理，要求导入后台数据库的知识图谱规模总量大于一千万条（10 million）三元组；知识图谱数据库不限，推荐使用Neo4j（管理RDF需要安装neosemantics插件）等。
4. 前台开发要求：前台用户界面要求是基于Web浏览器的，具体前台Web界面技术不限，例如可以是基于JavaScript的Node.js（如在知识图谱可视化课程中演示的访问Neo4j中知识图谱的Node.js项目）、基于Java的Spring Boot等。
5. 系统功能要求：
   1. 用户界面支持有实际意义的知识图谱查询，实现的知识图谱查询数量为3个或以上。
   2. 用户界面支持知识图谱可视化（基于G6或其他可视化组件工具），能够实现力导向布局。
   3. 用户界面支持3种或以上的知识图谱数据挖掘算法（例如，中心性、路径搜索、社区检测、相似性等类算法；推荐使用课程中介绍的Neo4j Graph Data Science组件），并给出算法的执行结果。
6. 作业提交要求
7. 以小组为团队完成项目系统开发。
8. 使用PPT验收汇报10分钟，录制讲解PPT和展示所开发系统的录屏视频。
9. 每名组员都应进行汇报讲解与展示，清楚汇报每人所完成的任务分工与工作量情况（PPT中也应有文字清楚标明）。
10. 完成项目报告，按照给定模板要求撰写。
11. 将以下材料打包为zip格式压缩包进行提交，命名为“组长学号.zip”：
    * + 文件夹：数据，包括项目用到的COVID-19知识图谱数据集中的数据文件；
      + 文件夹：系统源代码（在项目报告中写明部署过程，基于提供的数据和源代码系统能够运行，不依赖于其他数据、代码或工具），包括项目的所有源代码文件；
      + 文件：汇报PPT文件，幻灯片设置为16:9；
      + 文件：录屏视频文件，要求使用腾讯会议录屏功能，时长不超过10分钟，录屏视频文件格式为mp4；
      + 文件：包含小组所有成员学号姓名的文本文件，格式为txt。
12. 参考查询示例

以下是COVID-19知识图谱上的一些具有意义的SPARQL查询示例。更多的查询示例可以通过COVID-19知识图谱网站获取。

1. 查询String数据集中蛋白ACE2的蛋白相互作用网络PPI

match (Protein1\_ID: Resource.obo {obo: “RO\_0002160”}, Protein2\_ID: Resource.obo, Protein1\_Name, Protein2\_Name: Resource.rdf\_\_label, Protein2\_Name: Resource.rdf\_\_label)

where

###########

# SPPIQ1: String Human Protein Protein Interactions by Protein ID

###########

PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>

PREFIX obo: <http://purl.obolibrary.org/obo/>

PREFIX owl: <http://www.w3.org/2002/07/owl#>

PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>

PREFIX ensembl: <http://www.ensembl.org/id/>

PREFIX sio: <http://semanticscience.org/resource/>

PREFIX swo: <http://www.ebi.ac.uk/swo/>

SELECT ?Protein1\_ID ?Protein2\_ID ?Protein1\_Name ?Protein2\_Name

FROM <https://research.bioinformatics.udel.edu/covid-19-kg/string-human>

WHERE {

?Protein1 a owl:Class ;

obo:id ?Protein1\_ID ;

rdfs:label ?Protein1\_Name ;

rdfs:subClassOf [

a owl:Restriction ;

owl:onProperty obo:RO\_0002160 ;

owl:someValuesFrom ?Protein1\_Taxon;

] .

?Protein2 a owl:Class ;

obo:id ?Protein2\_ID ;

rdfs:label ?Protein2\_Name ;

rdfs:subClassOf [

a owl:Restriction ;

owl:onProperty obo:RO\_0002160 ;

owl:someValuesFrom ?Protein2\_Taxon;

] .

[

a owl:Axiom ;

owl:annotatedProperty sio:SIO\_000701 ;

owl:annotatedSource ?Protein1 ;

owl:annotatedTarget ?Protein2 ;

swo:SWO\_0000425 ?Combined\_score ;

rdfs:comment "combined\_score"

] .

FILTER(?Protein1\_ID = 'ENSP00000389326')

}

ORDER BY DESC(?Combined\_score)

1. 查询String数据集中一组给定蛋白的蛋白相互作用网络PPI

###########

# SPPIQ3: Interactions within a small set of proteins

###########

PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>

PREFIX obo: <http://purl.obolibrary.org/obo/>

PREFIX owl: <http://www.w3.org/2002/07/owl#>

PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>

PREFIX ensembl: <http://www.ensembl.org/id/>

PREFIX sio: <http://semanticscience.org/resource/>

PREFIX swo: <http://www.ebi.ac.uk/swo/>

SELECT ?Protein1\_ID ?Protein2\_ID ?Protein1\_Name ?Protein2\_Name

FROM <https://research.bioinformatics.udel.edu/covid-19-kg/string-human>

WHERE {

values ?Protein1\_ID { "ENSP00000386559" "ENSP00000276927" "ENSP00000011653" "ENSP00000389326" "ENSP00000229135" "ENSP00000369213" "ENSP00000226574" "ENSP00000385675" }

values ?Protein2\_ID { "ENSP00000386559" "ENSP00000276927" "ENSP00000011653" "ENSP00000389326" "ENSP00000229135" "ENSP00000369213" "ENSP00000226574" "ENSP00000385675" }

?Protein1 a owl:Class ;

obo:id ?Protein1\_ID ;

rdfs:label ?Protein1\_Name ;

rdfs:subClassOf [

a owl:Restriction ;

owl:onProperty obo:RO\_0002160 ;

owl:someValuesFrom ?Protein1\_Taxon;

] .

?Protein2 a owl:Class ;

obo:id ?Protein2\_ID ;

rdfs:label ?Protein2\_Name ;

rdfs:subClassOf [

a owl:Restriction ;

owl:onProperty obo:RO\_0002160 ;

owl:someValuesFrom ?Protein2\_Taxon;

] .

[

a owl:Axiom ;

owl:annotatedProperty sio:SIO\_000701 ;

owl:annotatedSource ?Protein1 ;

owl:annotatedTarget ?Protein2 ;

swo:SWO\_0000425 ?Combined\_score ;

rdfs:comment "combined\_score"

] .

FILTER(?Protein1\_ID != ?Protein2\_ID)

}

ORDER BY ASC(?Protein1\_ID) DESC(?Combined\_score)

1. 查询COVID19KG知识图谱中病毒蛋白和人类蛋白的相互作用网络

###########

# PPPQ2: Get SARS-CoV-2 protein interaction map

###########

PREFIX sars2: <https://research.bioinformatics.udel.edu/covid-19-kg/sars2\_ppi#>

SELECT ?viral\_protein ?human\_protein ?viral\_protein ?human\_gene\_name

FROM <https://research.bioinformatics.udel.edu/covid-19-kg/sars2\_ppi>

WHERE

{

[

sars2:viral\_protein ?viral\_protein ;

sars2:human\_protein ?human\_protein ;

sars2:human\_gene\_name ?human\_gene\_name ;

sars2:uniprot\_id ?uniprot\_id ;

sars2:uniprot\_description ?uniprot\_description ;

sars2:pdb ?pdb ;

sars2:disease ?disease ;

sars2:mist ?mist;

sars2:saint\_bfdr ?saint\_bfdr ;

sars2:avg\_spec ?avg\_spec ;

sars2:fold\_change ?fold\_change ;

] .

}

1. 查询DrugBank知识图谱中潜在药物和靶点的关系网络

###########

# DBQ3: Potential Drug Targets of COVID-19

###########

PREFIX drugbank: <https://research.bioinformatics.udel.edu/covid-19-kg/drugbank#>

SELECT DISTINCT ?drugbank\_id ?target\_id ?drugbank\_name ?target\_name

FROM <https://research.bioinformatics.udel.edu/covid-19-kg/drugbank>

WHERE {

values ?drugbank\_id { "DB00112" "DB00207" "DB00503" "DB00608" "DB00959" "DB01050" "DB01234" "DB01264" "DB01601" "DB05941" "DB06273" "DB08868" "DB11574" "DB11676" "DB12466" "DB12668" "DB13609" "DB14761" "DB14999" "DB15622" "DB15623" "DB15654" "DB15655" "DB15656" "DB15660" "DB15661" "DB15686" "DB15687" "DB15688" "DB15691" "DB15692" "DB15693"}

[

drugbank:drugbank-id ?drugbank\_id ;

drugbank:name ?drugbank\_name ;

drugbank:description ?drugbank\_description ;

drugbank:groups

[ drugbank:group ?group ] ;

drugbank:targets

[ drugbank:target

[ drugbank:id ?target\_id ;

drugbank:known-action ?known\_action ;

drugbank:name ?target\_name ;

drugbank:organism ?organism ;

drugbank:polypeptide

[

drugbank:gene-name ?gene

]

]

]

]

}

ORDER BY ?drugbank\_name

1. 查询miRTex标注工具标注的PubMed数据集中Covid-19相关的实体关系网络

###########

# CIMMQ8: Relations of CORD-19 Dataset (PubMed) Annotated By miRTex

###########

PREFIX mirtex: <https://research.bioinformatics.udel.edu/itextmine/cord19/medline/mirtex#>

SELECT DISTINCT ?entity\_duid1 ?entity\_text1 ?entity\_duid2 ?entity\_text2

FROM <https://research.bioinformatics.udel.edu/itextmine/cord19/medline/mirtex>

WHERE {

[ mirtex:docId ?docId ;

mirtex:entity

[ ?entity1

[ mirtex:duid

?entity\_duid1 ;

mirtex:entityText

?entity\_text1 ;

]

];

mirtex:entity

[ ?entity2

[ mirtex:duid

?entity\_duid2 ;

mirtex:entityText

?entity\_text2 ;

]

];

mirtex:relation

[ ?relation

[ mirtex:argument

[ mirtex:entity\_duid ?entity\_duid1 ;

mirtex:role ?role1

] ;

mirtex:argument

[ mirtex:entity\_duid ?entity\_duid2 ;

mirtex:role ?role2

] ;

mirtex:duid ?relation\_duid ;

mirtex:relationType ?relationType ;

mirtex:source ?source

]

]

]

}

ORDER BY ?docId

1. 查询eFIP标注工具标注的PubMed数据集中Covid-19相关的实体关系网络

###########

# CIMEQ8: Relations of CORD-19 Dataset (PubMed) Annotated By eFIP

###########

PREFIX efip: <https://research.bioinformatics.udel.edu/itextmine/cord19/medline/efip#>

SELECT DISTINCT ?entity\_duid1 ?entity\_text1 ?entity\_duid2 ?entity\_text2

FROM <https://research.bioinformatics.udel.edu/itextmine/cord19/medline/efip>

WHERE {

[ efip:docId ?docId ;

efip:entity

[ ?entity1

[ efip:duid

?entity\_duid1 ;

efip:entityText

?entity\_text1 ;

]

];

efip:entity

[ ?entity2

[ efip:duid

?entity\_duid2 ;

efip:entityText

?entity\_text2 ;

]

];

efip:relation

[ ?relation

[ efip:argument

[ efip:entity\_duid ?entity\_duid1 ;

efip:role ?role1

] ;

efip:argument

[ efip:entity\_duid ?entity\_duid2 ;

efip:role ?role2

] ;

efip:duid ?relation\_duid ;

efip:relationType ?relationType ;

efip:source ?source

]

]

]

}

ORDER BY ?docId