

گزارش پروژه نهایی وب معنایی

موضوع گراف دانش: Covid19

اعضای گروه: امیرحسین صمدپور، محمد کاهانی، ابوالفضل مهاجری

مقدمه

این گزارش شامل مراحل ایجاد گراف دانش مربوط به Covid-19 است. برای انجام این پروژه پس از بررسی دامنه، ابتدا به جمع‌آوری و مرتب‌سازی داده‌های مورد نیاز پرداختیم و پس از آن، بعد از ایجاد ارکان اصلی گراف دانش (Class, ObjectProperty و DataProperty)، داده‌های جمع‌آوری شده را به کمک ابزار Protege به سیستم اضافه کردیم. در نهایت کویری‌های SPARQL و scriptهای ساده‌ای برای تست و بررسی گراف دانش خود طراحی کردیم که در ادامه با جزییات این مراحل توضیح داده خواهند شد.

داده‌ها

موضوع Covid-19 از آنجایی که موضوع روز دنیاست و پژوهش‌های زیادی در این زمینه انجام شده است، دیتاست‌های زیادی نیز روزانه آپدیت و آماده می‌شود که ما برای پروژه برخی از آن‌ها را انتخاب کردیم. نهایتاً داده‌ها در سطح Country جمع‌آوری شده‌اند. (دیتاست‌های ذخیره شده را از داخل پوشه datasets می‌توانید مشاهده کنید.)

Countries & Statistics

برای جمع‌آوری داده کشورها از دیتاست آماری [GoogleCloud](#) استفاده کردیم. شمایی از دیتاست در شکل زیر قابل مشاهده است.

CountryName	DBPedia Country Links	TCases	TDeaths	TRecovered	ActiveCases	TTested

Vaccine

برای جمع‌آوری واکسن‌های تولید شده برای درمان Covid-19 روی صفحات Wikipedia کراال کرده و در نهایت جدولی با شمای زیر استخراج کردیم. لازم به ذکر است واکسن‌های استخراج شده به دو دسته قابل استفاده و در حال آزمایش تقسیم می‌شوند.

VaccineName	VaccineCountry	VaccineTechnology
-------------	----------------	-------------------

Restrictions

یکی از داده‌های جالب قابل جمع‌آوری مربوط به اخبار و اقدامات دولت کشورها در قبال بیماری کرونا بود که پس از جمع‌آوری اطلاعات به تفکیک سه موضوع تعطیلی مدارس، تعطیلی کار و تعطیلی پروازهای هواپیمایی اطلاعات را استخراج کردیم. در مورد تعطیلی پروازها دیتاستی که به تفکیک هواپیمایی محدودیت‌های اعمال شده را نشان می‌داد جمع‌آوری کردیم. شمای این داده‌ها در زیر قابل مشاهده است.

Countries	Date	School News	WorkPlace News
-----------	------	-------------	----------------

ارتباط با گراف‌های دانش خارجی

برای اتصال موجودیت‌ها و مفاهیم به گراف‌های خارجی مانند dbpedia از اسکریپت پایتون استفاده کردیم و روی صفحات wikipedia به ازای اسامی کشورها، واکسن‌ها و هواپیمایی‌ها آن را ران کردیم. هم‌چنین برای تعریف مفاهیم از schema.org استفاده کردیم. نمونه‌ای از تعریف مفهوم و خروجی اسکریپت را می‌توانید در زیر مشاهده کنید.

Country	Resources
USA	http://dbpedia.org/resource/United_States
India	http://dbpedia.org/resource/India
Brazil	http://dbpedia.org/resource/Brazil
Russia	http://dbpedia.org/resource/Russia
France	http://dbpedia.org/resource/France
Turkey	http://dbpedia.org/resource/Turkey
UK	http://dbpedia.org/resource/United_Kingdom
Argentina	http://dbpedia.org/resource/Argentina
Colombia	http://dbpedia.org/resource/Colombia
Italy	http://dbpedia.org/resource/Italy
Spain	http://dbpedia.org/resource/Spain
Germany	http://dbpedia.org/resource/Germany
Iran	http://dbpedia.org/resource/Iran
Poland	http://dbpedia.org/resource/Poland

تعریف کشور:

```
<owl:Class rdf:about="https://schema.org/Country">
  <rdfs:subClassOf rdf:resource="https://schema.org/Place"/>
</owl:Class>
```

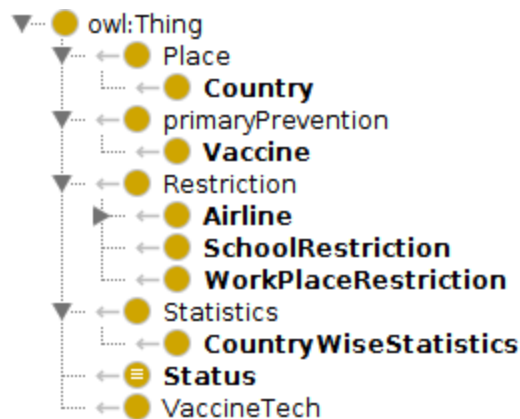
تعریف کشور ایران:

```
<owl:NamedIndividual
rdf:about="http://dbpedia.org/resource/Iran">
  <rdf:type rdf:resource="https://schema.org/Country"/>
  <rdfs:label
rdf:datatype="http://www.w3.org/2001/XMLSchema#string">Iran</rdf
s:label>
</owl:NamedIndividual>
```

گراف دانش

موجودیت‌ها و ارتباطات

پس از جمع‌آوری داده‌ها و تعریف دامنه پروژه گراف خود را طراحی کردیم که ساختار کلی آن از طریق فایل svg داخل پوشه doc قابل مشاهده است. به طور کلی کلاس‌هایمان شامل موارد زیر است:



نحوه وارد کردن داده به Protege

برای وارد کردن داده‌ها ابزار Protege از پلاگینی به نام Cellfie استفاده کردیم که به عنوان ورودی فایل اکسل دریافت می‌کند و به کمک آن می‌توان با تعریف قوانین مختلف با توجه ستون‌ها داده‌ها را برای مفهوم خاص به صورت یکپارچه و نیمه‌خودکار اضافه کرد. در ادامه نمونه‌ای از وارد کردن داده‌های هواپیمایی‌ها قابل مشاهده است:

```

Individual: @I*(mm:printf("\'%sAirlineRestriction_%s\'", @I*, @A*))
types: AirlineRestriction
Facts: airlineRestrictionOf @I*,
      Notes @K*
      RestrictionDate @G*(xsd:date)
  
```

پرس وجو روی گراف دانش

برای ارزیابی و بررسی بیشتر گراف دانش پرس و جوهای زیر را آماده کردیم.

پرس و جوی اول

تعداد کل واکسن‌های ساخته شده (چند نوع):

PREFIX :

<<http://www.semanticweb.org/team1/ontologies/2021/covid19/>>

PREFIX sch: <<https://schema.org/>>

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

SELECT DISTINCT COUNT(?vaccine)

WHERE {

 ?vaccine a :Vaccine

Query
Saved Queries

Default Graph IRI

Query

PREFIX : <http://www.semanticweb.org/team1/ontologies/2021/covid19/>
PREFIX sch: <https://schema.org/>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>

SELECT DISTINCT COUNT(?vaccine)
WHERE {
?vaccine a :Vaccine
}

Execute
Save
Load
Clear

SPARQL | HTML5 table

callret-0

114

پرس و جوی دوم

لیست هواپیمایی‌های دارای محدودیت بخاطر Covid19:

```

PREFIX :
<http://www.semanticweb.org/team1/ontologies/2021/covid19/>
PREFIX sch: <https://schema.org/>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>

SELECT DISTINCT ?airline

```

```

WHERE {
    ?restriction a :AirlineRestriction;
                :airlineRestrictionOf ?airline;
                :RestrictionDate ?date.
}

```

Query
Saved Queries

Default Graph IRI
http://localhost:8890/DAV

Query

```

PREFIX : <http://www.semanticweb.org/team1/ontologies/2021/covid19/>
PREFIX sch: <https://schema.org/>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>

SELECT DISTINCT ?airline
WHERE {
    ?restriction a :AirlineRestriction;
                :airlineRestrictionOf ?airline;
                :RestrictionDate ?date.
}

```

Execute
Save
Load
Clear

SPARQL | HTML5 table

airline
http://dbpedia.org/resource/EVA_Air
http://dbpedia.org/resource/ANA_&_JP_Express
http://dbpedia.org/resource/ASL_Airlines_France
http://dbpedia.org/resource/Malév_Hungarian_Airlines
http://dbpedia.org/resource/Aer_Lingus_Regional
http://dbpedia.org/resource/Aeroméxico_Connect
http://dbpedia.org/resource/Aeroflot
http://dbpedia.org/resource/Aerolíneas_Argentinas
http://dbpedia.org/resource/Air_Algerie
http://dbpedia.org/resource/Air_Antilles
http://dbpedia.org/resource/Air_Antwerp
http://dbpedia.org/resource/Air_Arabia
http://dbpedia.org/resource/AirAsia
http://dbpedia.org/resource/AirAsia_India
http://dbpedia.org/resource/Air_Astana
http://dbpedia.org/resource/Air_Austral
http://dbpedia.org/resource/ASL_Airlines_Belgium
http://dbpedia.org/resource/Air_Burkina

لیست کشورهایی که برای مدارس محدودیتی ایجاد نکردند (البته تاریخ اهمیت دارد):

PREFIX :

<<http://www.semanticweb.org/team1/ontologies/2021/covid19/>>

PREFIX sch: <<https://schema.org/>>

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

SELECT ?restriction ?note

WHERE {

 ?restriction a :SchoolRestriction.

 ?restriction :Notes ?note.

 FILTER (!regex(?note, "no restriction applied", "i"))

}

Query
Saved Queries

Default Graph IRI
http://localhost:8890/DAV

Query

```

PREFIX : <http://www.semanticweb.org/team1/ontologies/2021/covid19/>
PREFIX sch: <https://schema.org/>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>

SELECT DISTINCT ?restriction
WHERE {
  ?restriction a :SchoolRestriction.
  ?restriction :Notes ?note.
  FILTER (!regex(?note, "no restriction applied", "i"))
}

```

Execute
Save
Load
Clear

SPARQL | HTML5 table

restriction
http://www.semanticweb.org/team1/ontologies/2021/covid19/AfghanistanSchoolRestriction_1
http://www.semanticweb.org/team1/ontologies/2021/covid19/AfghanistanSchoolRestriction_10161
http://www.semanticweb.org/team1/ontologies/2021/covid19/AfghanistanSchoolRestriction_10348
http://www.semanticweb.org/team1/ontologies/2021/covid19/AfghanistanSchoolRestriction_1041
http://www.semanticweb.org/team1/ontologies/2021/covid19/AfghanistanSchoolRestriction_105
http://www.semanticweb.org/team1/ontologies/2021/covid19/AfghanistanSchoolRestriction_1145
http://www.semanticweb.org/team1/ontologies/2021/covid19/AfghanistanSchoolRestriction_1249
http://www.semanticweb.org/team1/ontologies/2021/covid19/AfghanistanSchoolRestriction_1353
http://www.semanticweb.org/team1/ontologies/2021/covid19/AfghanistanSchoolRestriction_1457
http://www.semanticweb.org/team1/ontologies/2021/covid19/AfghanistanSchoolRestriction_1561
http://www.semanticweb.org/team1/ontologies/2021/covid19/AfghanistanSchoolRestriction_1665
http://www.semanticweb.org/team1/ontologies/2021/covid19/AfghanistanSchoolRestriction_1769
http://www.semanticweb.org/team1/ontologies/2021/covid19/AfghanistanSchoolRestriction_1873
http://www.semanticweb.org/team1/ontologies/2021/covid19/AfghanistanSchoolRestriction_1977

پرس و جوی چهارم

کشورهایی که بیش تر از صد هزار مرگ به دلیل کرونا داشته اند:

PREFIX :

<<http://www.semanticweb.org/team1/ontologies/2021/covid19/>>

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

```
SELECT ?country ?deaths
WHERE {
    ?country a :CountryWiseStatistics.
    ?country :Death ?deaths.
    FILTER (?deaths > 100000).
}
```

Query Saved Queries

Default Graph IRI

Query

```
PREFIX : <http://www.semanticweb.org/team1/ontologies/2021/covid19/>
PREFIX sch: <https://schema.org/>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>

SELECT ?country ?deaths
WHERE {
  ?country a :CountryWiseStatistics.
  ?country :Death ?deaths.
  FILTER (?deaths > 100000).
}
```

Execute

Save

Load

Clear

SPARQL | HTML5 table

country	deaths
http://www.semanticweb.org/team1/ontologies/2021/covid19/ArgentinaStat	100695
http://www.semanticweb.org/team1/ontologies/2021/covid19/BrazilStat	539050
http://www.semanticweb.org/team1/ontologies/2021/covid19/ColombiaStat	114833
http://www.semanticweb.org/team1/ontologies/2021/covid19/FranceStat	111451
http://www.semanticweb.org/team1/ontologies/2021/covid19/IndiaStat	413123
http://www.semanticweb.org/team1/ontologies/2021/covid19/ItalyStat	127851
http://www.semanticweb.org/team1/ontologies/2021/covid19/MexicoStat	235740
http://www.semanticweb.org/team1/ontologies/2021/covid19/PeruStat	194845
http://www.semanticweb.org/team1/ontologies/2021/covid19/RussiaStat	146868
http://www.semanticweb.org/team1/ontologies/2021/covid19/UKStat	128642
http://www.semanticweb.org/team1/ontologies/2021/covid19/USASat	624297

پرس و جوی پنجم

لیست تمام واکسن‌هایی که ایران ساخته است:

PREFIX :

<<http://www.semanticweb.org/team1/ontologies/2021/covid19/>>

PREFIX sch: <<https://schema.org/>>

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

SELECT ?x ?y

WHERE {

 ?x a :Vaccine;

 :vaccineOf ?y.

 ?y rdfs:label "Iran"^^xsd:string

}

Query Saved Queries

Default Graph IRI

Query

```
PREFIX : <http://www.semanticweb.org/team1/ontologies/2021/covid19/>
PREFIX sch: <https://schema.org/>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>

SELECT ?x ?y
WHERE {
  ?x a :Vaccine;
      :vaccineOf ?y.
  ?y rdfs:label "Iran"^^xsd:string
}
```

Execute

Save

Load

Clear

SPARQL | HTML5 table

x	y
http://dbpedia.org/resource/COVIranBarakat	http://dbpedia.org/resource/Iran
http://dbpedia.org/resource/FAKHRAVAC	http://dbpedia.org/resource/Iran
http://dbpedia.org/resource/Noora	http://dbpedia.org/resource/Iran
http://dbpedia.org/resource/RaziCovPars	http://dbpedia.org/resource/Iran