SPARQL QUERIES in STARDOG 5.2.1

1. List the tobacco users who are CC patients.

select ?cc\_diag ?patient ?tu\_role

where

{{?cc\_diag a icd9cm:MM\_CLASS\_7355.}union{?cc\_diag a icd9cm:MM\_CLASS\_6577.}union{?cc\_diag a icd9cm:MM\_CLASS\_6581.}union{?cc\_diag a icd9cm:MM\_CLASS\_6578.}union{?cc\_diag a icd9cm:MM\_CLASS\_6579.}union{?cc\_diag a icd9cm:MM\_CLASS\_6580.}

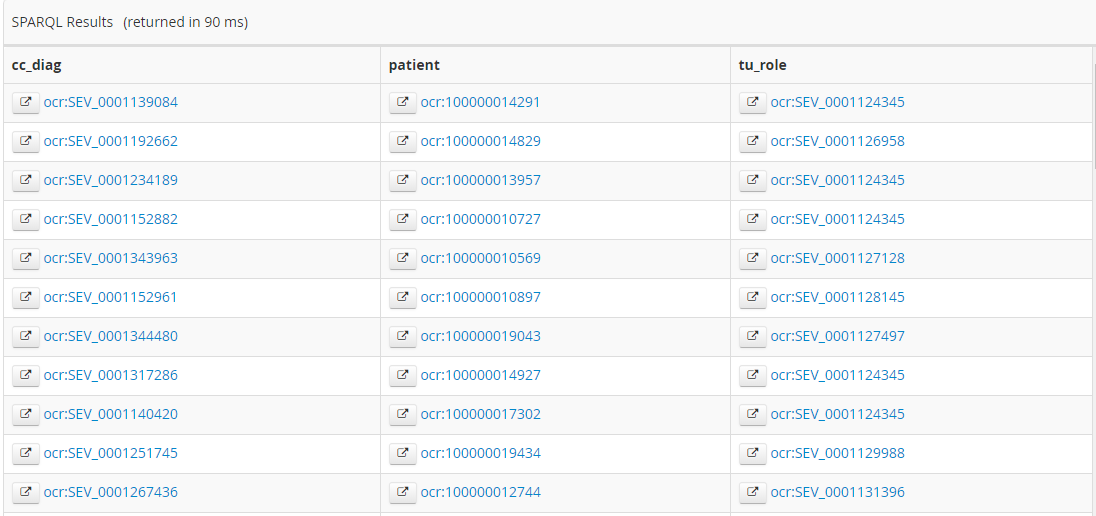
{?tu\_role a ocr:OCRSEV\_00000545.

?tu\_role obo:RO\_0000052 ?patient.

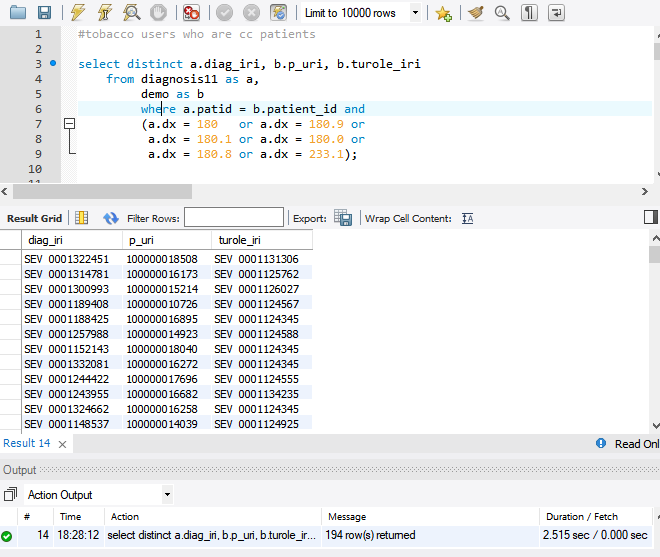
?cc\_diag obo:IAO\_0000136 ?patient.

?patient a obo:NCBITaxon\_9606.}}





SQL validation:



2. List the CC patients who have HIV:

select ?cc\_diag ?patient ?hiv\_diag

where

{{{?cc\_diag a icd9cm:MM\_CLASS\_7355.}union{?cc\_diag a icd9cm:MM\_CLASS\_6577.}union{?cc\_diag a icd9cm:MM\_CLASS\_6581.}union{?cc\_diag a icd9cm:MM\_CLASS\_6578.}union{?cc\_diag a icd9cm:MM\_CLASS\_6579.}union{?cc\_diag a icd9cm:MM\_CLASS\_6580.}}

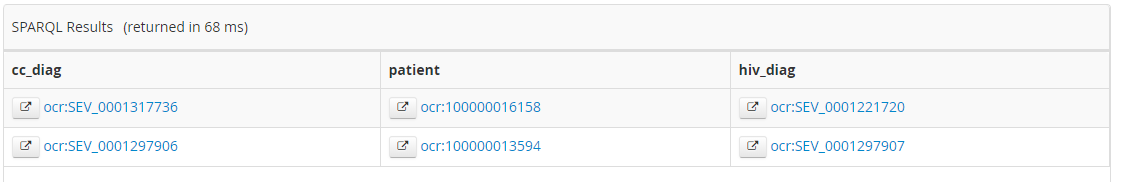
{?hiv\_diag a icd9cm:MM\_CLASS\_5563.

?hiv\_diag obo:IAO\_0000136 ?patient.

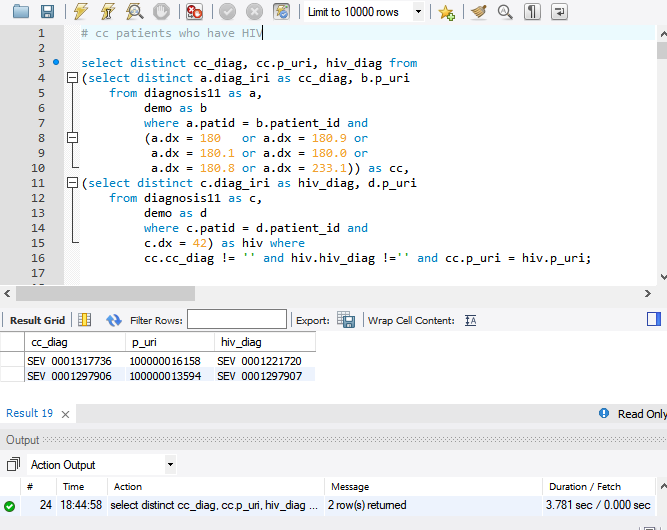
?cc\_diag obo:IAO\_0000136 ?patient.

?patient a obo:NCBITaxon\_9606.}}





SQL Validation



3. Calculate the age of all CC patients.

select (?pat as ?patient)

((xsd:integer(strafter(strafter(?diag\_date, '/'), '/')))-(xsd:integer(strafter(strafter(?pat\_date, '/'), '/'))) as ?age)

where

{

{{?cc\_diag a icd9cm:MM\_CLASS\_7355.}union{?cc\_diag a icd9cm:MM\_CLASS\_6577.}union{?cc\_diag a icd9cm:MM\_CLASS\_6581.}union{?cc\_diag a icd9cm:MM\_CLASS\_6578.}union{?cc\_diag a icd9cm:MM\_CLASS\_6579.}union{?cc\_diag a icd9cm:MM\_CLASS\_6580.}}

{?cc\_diag obo:OBI\_0000312 ?process.

?process obo:BFO\_0000155 ?interval.

?interval obo:RO\_0002091 ?date\_iri.

?date\_iri rdfs:label ?diag\_date.}

{?pat a obo:NCBITaxon\_9606.

?cc\_diag obo:IAO\_0000136 ?pat.

?pat obo:RO\_0000056 ?neo.

?neo obo:BFO\_0000155 ?tr2.}

optional{?birth a obo:UBERON\_0035946.

?birth obo:BFO\_0000139 ?neo.

?birth obo:BFO\_0000155 ?tr1.

{?tr1 obo:BFO\_0000139 ?tr2.

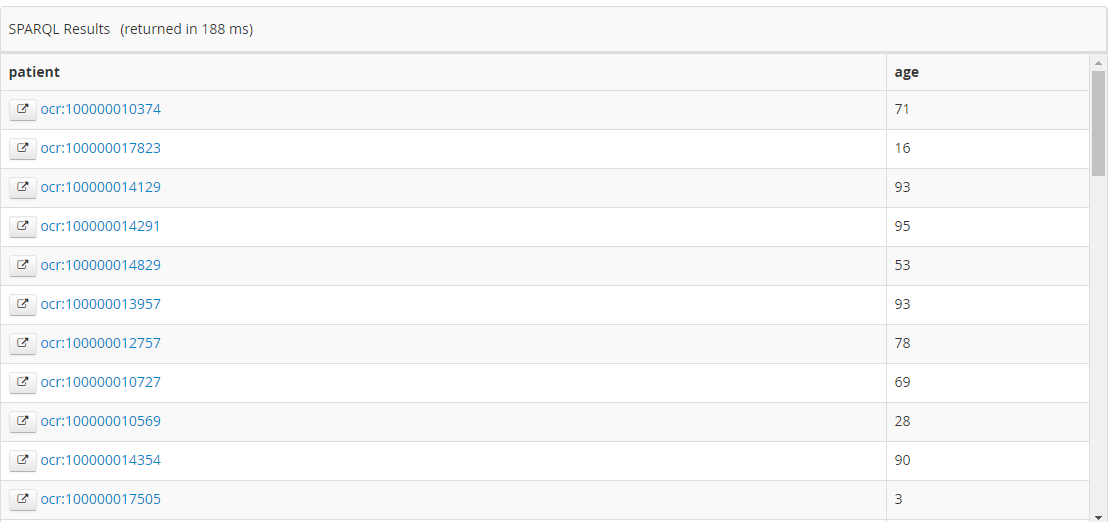
?tr1 obo:BFO\_0000139 ?tr3.

?tr3 rdfs:label ?pat\_date.}

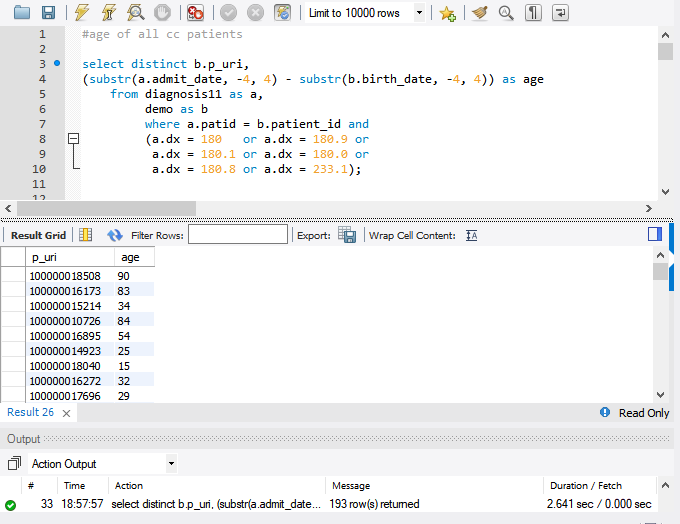
}

}

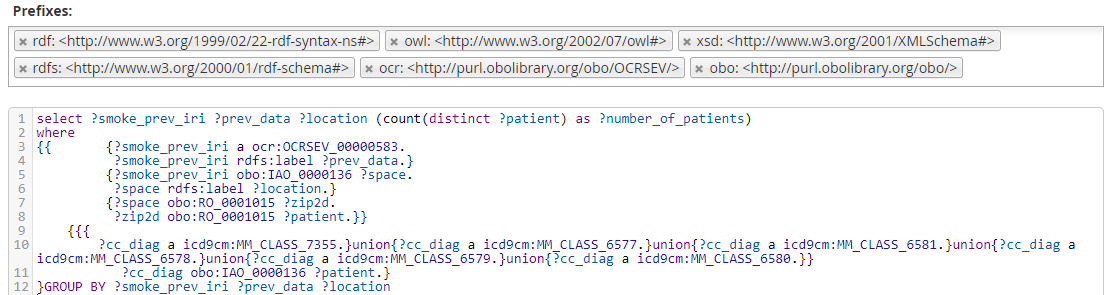




SQL Validation:



4. List the smoking prevalence within MMSAs along with the number of CC patients within each MMSA.



select ?smoke\_prev\_iri ?prev\_data ?location (count(distinct ?patient) as ?number\_of\_patients)

where

{{ {?smoke\_prev\_iri a ocr:OCRSEV\_00000583.

?smoke\_prev\_iri rdfs:label ?prev\_data.}

{?smoke\_prev\_iri obo:IAO\_0000136 ?space.

?space rdfs:label ?location.}

{?space obo:RO\_0001015 ?zip2d.

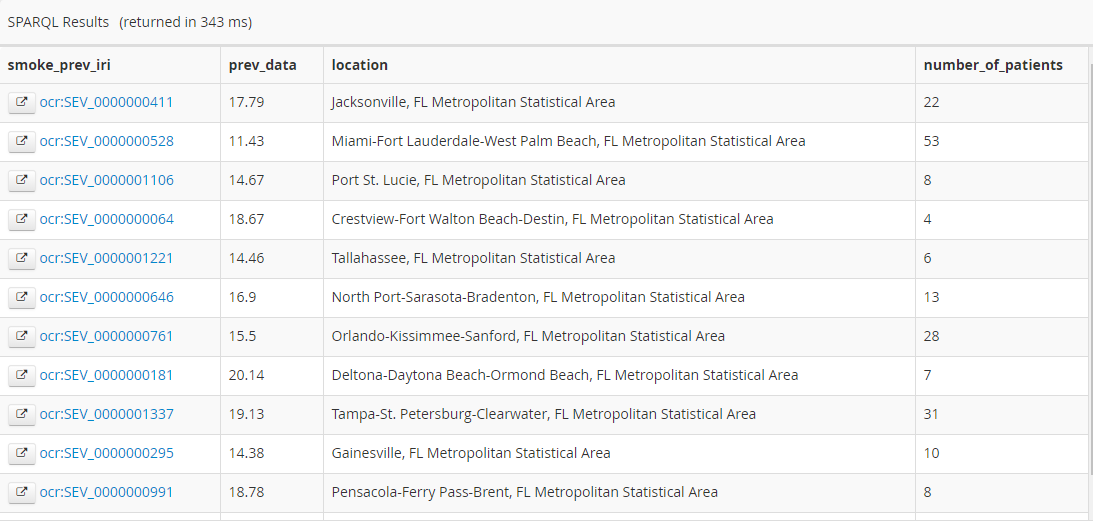
?zip2d obo:RO\_0001015 ?patient.}}

{{{

?cc\_diag a icd9cm:MM\_CLASS\_7355.}union{?cc\_diag a icd9cm:MM\_CLASS\_6577.}union{?cc\_diag a icd9cm:MM\_CLASS\_6581.}union{?cc\_diag a icd9cm:MM\_CLASS\_6578.}union{?cc\_diag a icd9cm:MM\_CLASS\_6579.}union{?cc\_diag a icd9cm:MM\_CLASS\_6580.}}

?cc\_diag obo:IAO\_0000136 ?patient.}

}GROUP BY ?smoke\_prev\_iri ?prev\_data ?location



5. List the drinking prevalence within MMSAs and the number of CC patients in each MMSA.

select ?drink\_prev\_iri ?prev\_data ?location (count(distinct ?patient) as ?number\_of\_patients)

where

{

{{?drink\_prev\_iri a ocr:OCRSEV\_00000601.

?drink\_prev\_iri rdfs:label ?prev\_data.}

{?drink\_prev\_iri obo:IAO\_0000136 ?space.

?space rdfs:label ?location.}

{?space obo:RO\_0001015 ?zip2d.

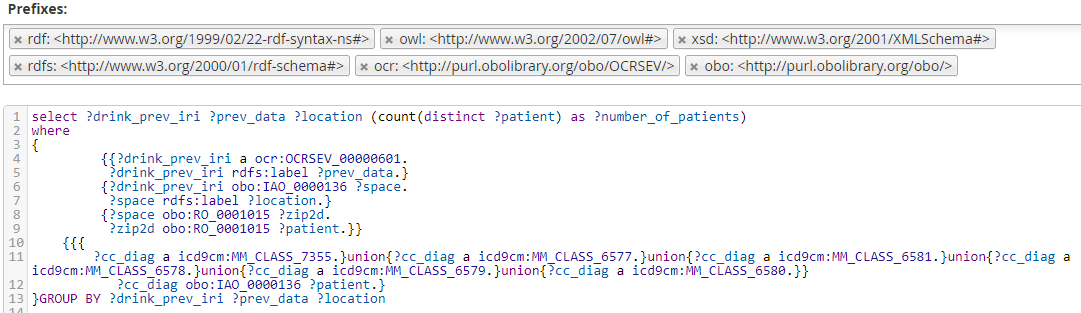
?zip2d obo:RO\_0001015 ?patient.}}

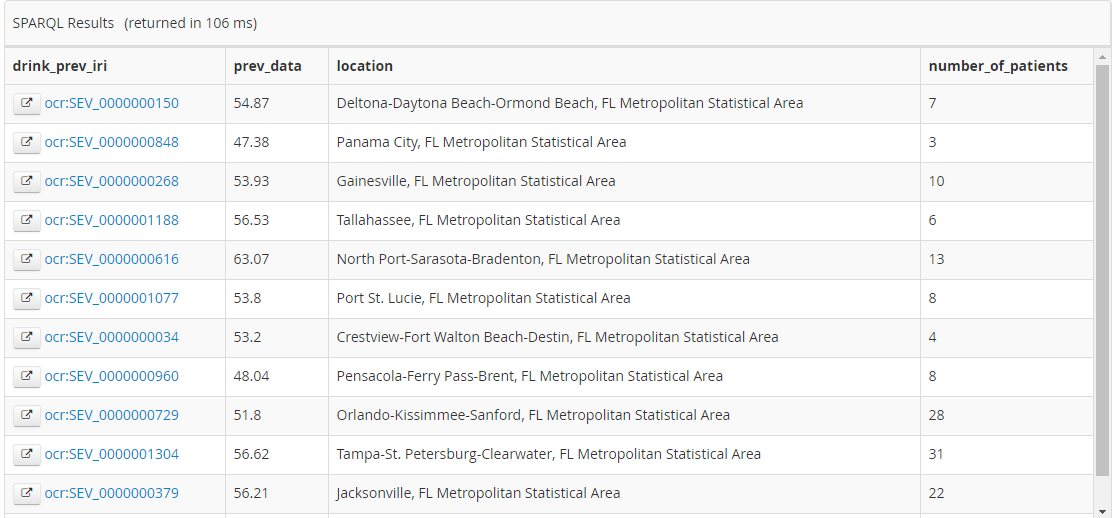
{{{

?cc\_diag a icd9cm:MM\_CLASS\_7355.}union{?cc\_diag a icd9cm:MM\_CLASS\_6577.}union{?cc\_diag a icd9cm:MM\_CLASS\_6581.}union{?cc\_diag a icd9cm:MM\_CLASS\_6578.}union{?cc\_diag a icd9cm:MM\_CLASS\_6579.}union{?cc\_diag a icd9cm:MM\_CLASS\_6580.}}

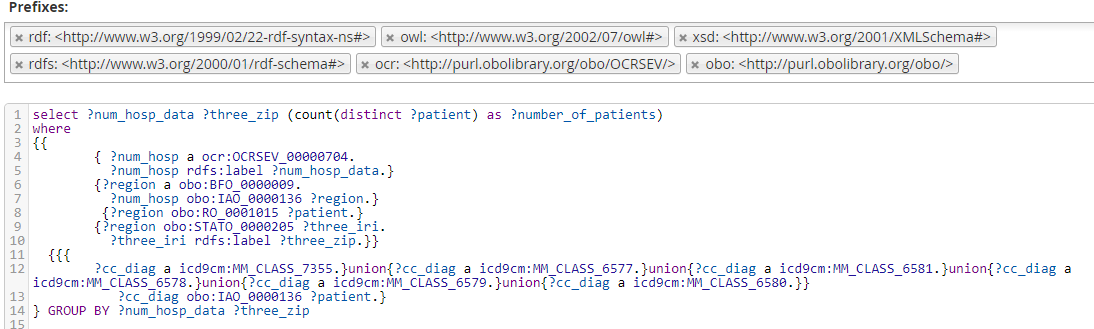
?cc\_diag obo:IAO\_0000136 ?patient.}

}GROUP BY ?drink\_prev\_iri ?prev\_data ?location





6. List the number of CC patients and the number of hospitals within a 3-digit zip.



select ?num\_hosp\_data ?three\_zip (count(distinct ?patient) as ?number\_of\_patients)

where

{{

{ ?num\_hosp a ocr:OCRSEV\_00000704.

?num\_hosp rdfs:label ?num\_hosp\_data.}

{?region a obo:BFO\_0000009.

?num\_hosp obo:IAO\_0000136 ?region.}

{?region obo:RO\_0001015 ?patient.}

{?region obo:STATO\_0000205 ?three\_iri.

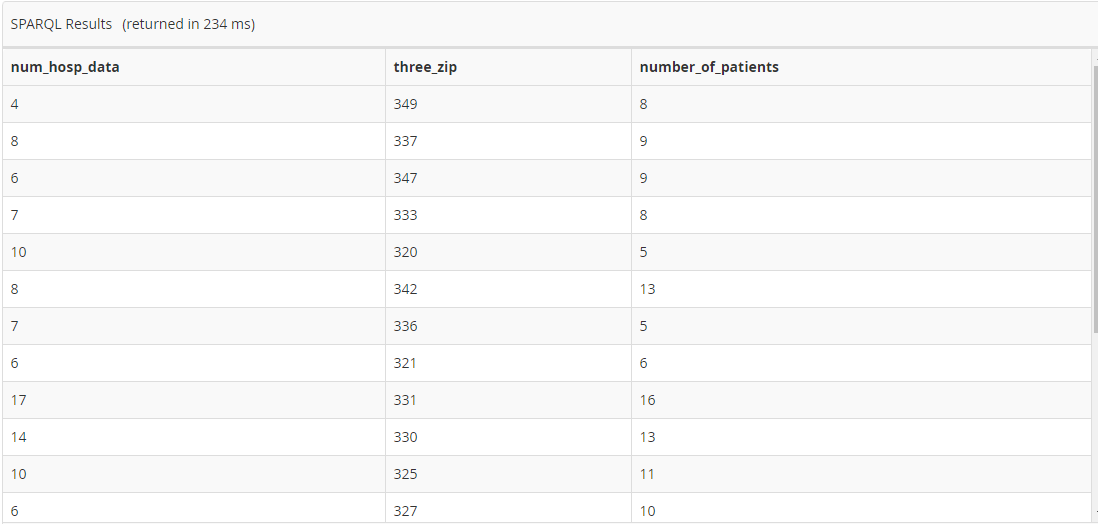
?three\_iri rdfs:label ?three\_zip.}}

{{{

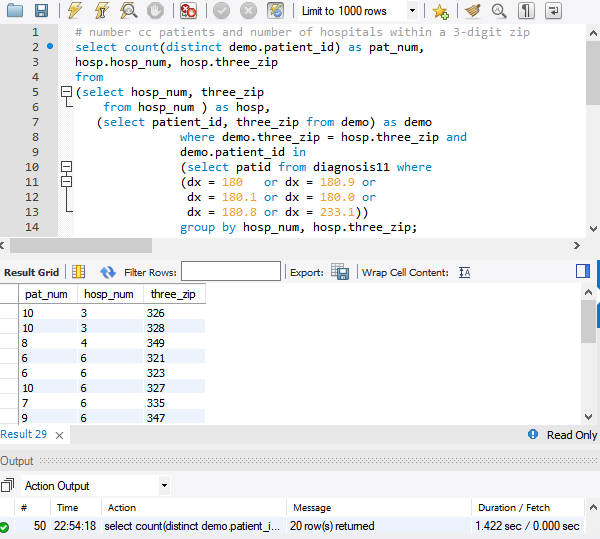
?cc\_diag a icd9cm:MM\_CLASS\_7355.}union{?cc\_diag a icd9cm:MM\_CLASS\_6577.}union{?cc\_diag a icd9cm:MM\_CLASS\_6581.}union{?cc\_diag a icd9cm:MM\_CLASS\_6578.}union{?cc\_diag a icd9cm:MM\_CLASS\_6579.}union{?cc\_diag a icd9cm:MM\_CLASS\_6580.}}

?cc\_diag obo:IAO\_0000136 ?patient.}

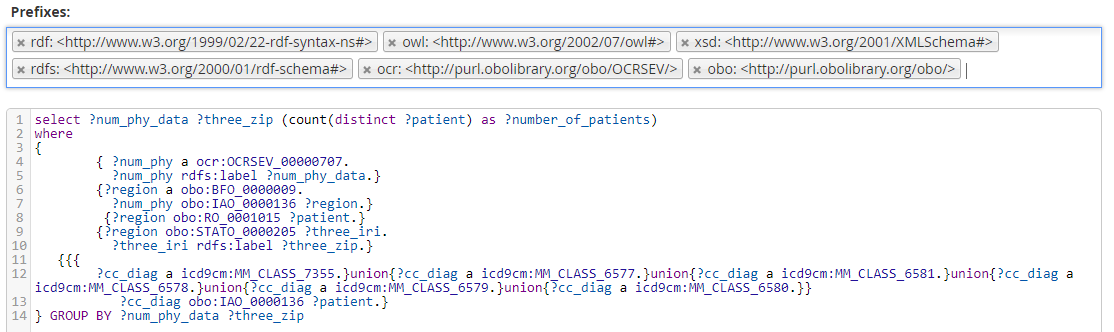
} GROUP BY ?num\_hosp\_data ?three\_zip



SQL Validation:



7. List the number of CC patients and number of physicians within a 3-digit zip.



select ?num\_phy\_data ?three\_zip (count(distinct ?patient) as ?number\_of\_patients)

where

{

{ ?num\_phy a ocr:OCRSEV\_00000707.

?num\_phy rdfs:label ?num\_phy\_data.}

{?region a obo:BFO\_0000009.

?num\_phy obo:IAO\_0000136 ?region.}

{?region obo:RO\_0001015 ?patient.}

{?region obo:STATO\_0000205 ?three\_iri.

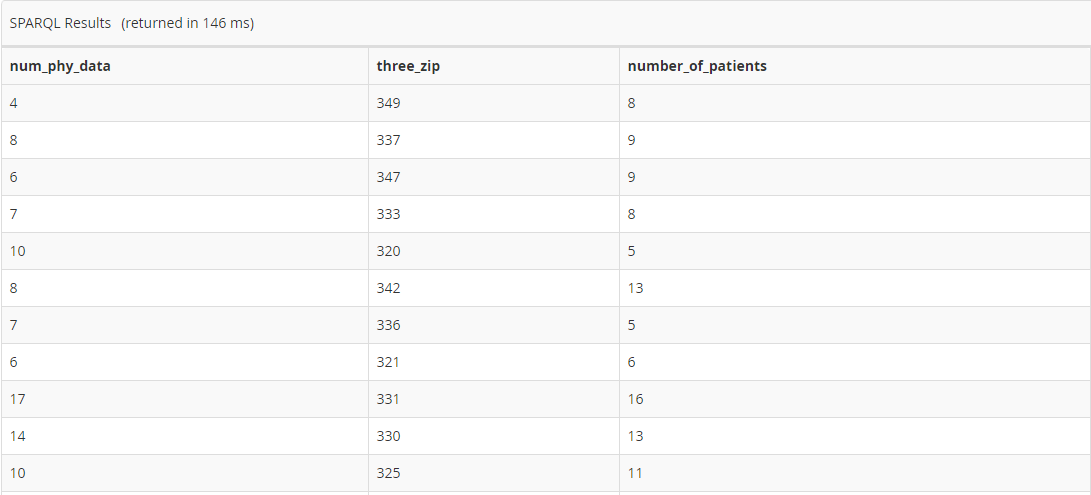
?three\_iri rdfs:label ?three\_zip.}

{{{

?cc\_diag a icd9cm:MM\_CLASS\_7355.}union{?cc\_diag a icd9cm:MM\_CLASS\_6577.}union{?cc\_diag a icd9cm:MM\_CLASS\_6581.}union{?cc\_diag a icd9cm:MM\_CLASS\_6578.}union{?cc\_diag a icd9cm:MM\_CLASS\_6579.}union{?cc\_diag a icd9cm:MM\_CLASS\_6580.}}

?cc\_diag obo:IAO\_0000136 ?patient.}

} GROUP BY ?num\_phy\_data ?three\_zip



8. List the number of cc patients per hospital along with hospital’s MSPB score which is < 1 (national median)

select distinct ?ccn ?mspb\_score (count(distinct ?patient) as ?pat\_num)

where

{{{?hosp a obo:OMRSE\_00000102.

?hosp obo:STATO\_0000205 ?ccn\_iri.

?hosp obo:OMRSE\_00000068 ?orgn\_iri.

?ccn\_iri rdfs:label ?ccn.

?mspb a ocr:OCRSEV\_00000740.

?mspb obo:IAO\_0000136 ?orgn\_iri.

?mspb rdfs:label ?mspb\_score.

{?gyn\_iri a obo:NCBITaxon\_9606.

?orgn\_iri obo:RO\_0002351 ?gyn\_iri.}

{?hce\_iri a obo:OGMS\_0000097.}

{?gyn\_iri obo:BFO\_0000056 ?hce\_iri.}

{?orgn\_iri obo:BFO\_0000056 ?hce\_iri.}

{?hce\_iri obo:BFO\_0000057 ?patient.}}

FILTER EXISTS{{{

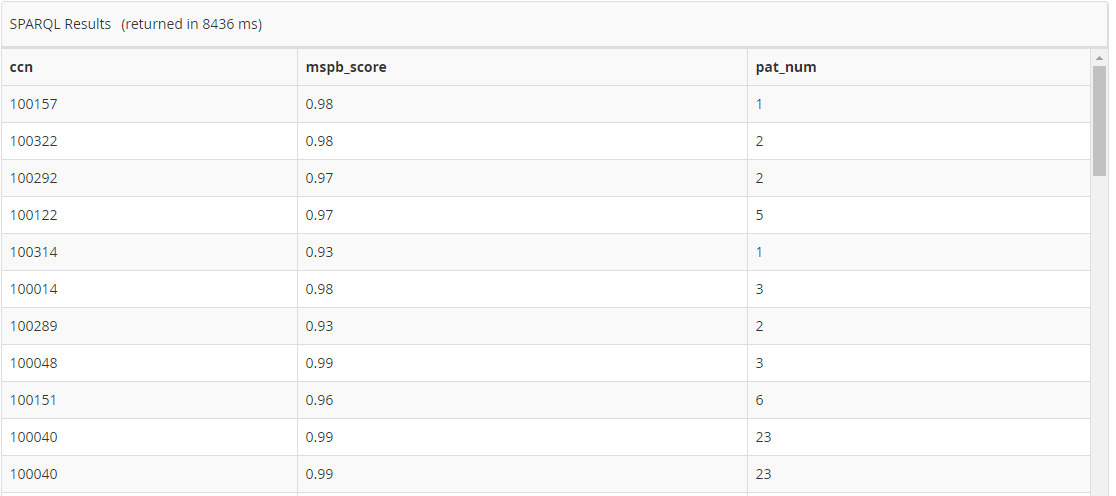
?cc\_diag a icd9cm:MM\_CLASS\_7355.}union{?cc\_diag a icd9cm:MM\_CLASS\_6577.}union{?cc\_diag a icd9cm:MM\_CLASS\_6581.}union{?cc\_diag a icd9cm:MM\_CLASS\_6578.}union{?cc\_diag a icd9cm:MM\_CLASS\_6579.}union{?cc\_diag a icd9cm:MM\_CLASS\_6580.}}

?cc\_diag obo:IAO\_0000136 ?patient.}}

FILTER((xsd:float(?mspb\_score)) < 1)

}GROUP BY ?ccn ?mspb\_score





9. What is the experience of every physician in years and how many CC patients has each physician encountered?

select ?npi ?grad\_year ((2016-(xsd:integer(?grad\_year))) as ?exp\_years) (count(distinct ?patient) as ?pat\_num)

where{

{{?hosp a obo:OMRSE\_00000102.

?hosp obo:OMRSE\_00000068 ?orgn\_iri.

?orgn\_iri obo:RO\_0002351 ?gyn\_iri.}

{?gyn\_iri a obo:NCBITaxon\_9606.}

{?gyn\_iri obo:BFO\_0000056 ?resi\_iri.}

{?gyn\_iri obo:STATO\_0000205 ?npi\_iri.}

{?npi\_iri rdfs:label ?npi.}

{?resi\_iri obo:RO\_0002093 ?year\_iri.

?year\_iri rdfs:label ?grad\_year. }

{?orgn\_iri obo:RO\_0002351 ?gyn\_iri.}

{?hce\_iri a obo:OGMS\_0000097.}

{?gyn\_iri obo:BFO\_0000056 ?hce\_iri.}

{?orgn\_iri obo:BFO\_0000056 ?hce\_iri.}

{?hce\_iri obo:BFO\_0000057 ?patient.}}

FILTER EXISTS{{{

?cc\_diag a icd9cm:MM\_CLASS\_7355.}union{?cc\_diag a icd9cm:MM\_CLASS\_6577.}union{?cc\_diag a icd9cm:MM\_CLASS\_6581.}union{?cc\_diag a icd9cm:MM\_CLASS\_6578.}union{?cc\_diag a icd9cm:MM\_CLASS\_6579.}union{?cc\_diag a icd9cm:MM\_CLASS\_6580.}}

?cc\_diag obo:IAO\_0000136 ?patient.}

} GROUP BY ?npi ?grad\_year





10. List the number of CC physicians per CC patient.

select distinct ?patient (count(distinct ?gyn\_iri) as ?gyn\_num)

where

{

{{?hosp a obo:OMRSE\_00000102.

?hosp obo:OMRSE\_00000068 ?orgn\_iri.

?orgn\_iri obo:RO\_0002351 ?gyn\_iri.}

{?gyn\_iri a obo:NCBITaxon\_9606.}

{?gyn\_iri obo:BFO\_0000056 ?resi\_iri.}

{?gyn\_iri obo:STATO\_0000205 ?npi\_iri.}

{?npi\_iri rdfs:label ?npi.}

{?resi\_iri obo:RO\_0002093 ?year\_iri.

?year\_iri rdfs:label ?grad\_year. }

{?orgn\_iri obo:RO\_0002351 ?gyn\_iri.}

{?hce\_iri a obo:OGMS\_0000097.}

{?gyn\_iri obo:BFO\_0000056 ?hce\_iri.}

{?orgn\_iri obo:BFO\_0000056 ?hce\_iri.}

{?hce\_iri obo:BFO\_0000057 ?patient.}}

FILTER EXISTS{{{

?cc\_diag a icd9cm:MM\_CLASS\_7355.}union{?cc\_diag a icd9cm:MM\_CLASS\_6577.}union{?cc\_diag a icd9cm:MM\_CLASS\_6581.}union{?cc\_diag a icd9cm:MM\_CLASS\_6578.}union{?cc\_diag a icd9cm:MM\_CLASS\_6579.}union{?cc\_diag a icd9cm:MM\_CLASS\_6580.}}

?cc\_diag obo:IAO\_0000136 ?patient.}

} GROUP BY ?patient





11. List the number of CC patients and average experience of physician per hospital.

select distinct ?ccn (FLOOR(AVG(2016-(xsd:integer(?grad\_year)))) as ?avg\_exp\_years) (count(distinct ?patient) as ?pat\_num)

where

{

{{?hosp a obo:OMRSE\_00000102.

?hosp obo:STATO\_0000205 ?ccn\_iri.

?hosp obo:OMRSE\_00000068 ?orgn\_iri.

?ccn\_iri rdfs:label ?ccn.

?orgn\_iri obo:RO\_0002351 ?gyn\_iri.}

{select distinct ?gyn\_iri ?npi ?grad\_year ?patient where{

{?gyn\_iri a obo:NCBITaxon\_9606.}

{?gyn\_iri obo:BFO\_0000056 ?resi\_iri.}

{?gyn\_iri obo:STATO\_0000205 ?npi\_iri.}

{?npi\_iri rdfs:label ?npi.}

{?resi\_iri obo:RO\_0002093 ?year\_iri.

?year\_iri rdfs:label ?grad\_year. }

{?orgn\_iri obo:RO\_0002351 ?gyn\_iri.}

{?hce\_iri a obo:OGMS\_0000097.}

{?gyn\_iri obo:BFO\_0000056 ?hce\_iri.}

{?orgn\_iri obo:BFO\_0000056 ?hce\_iri.}

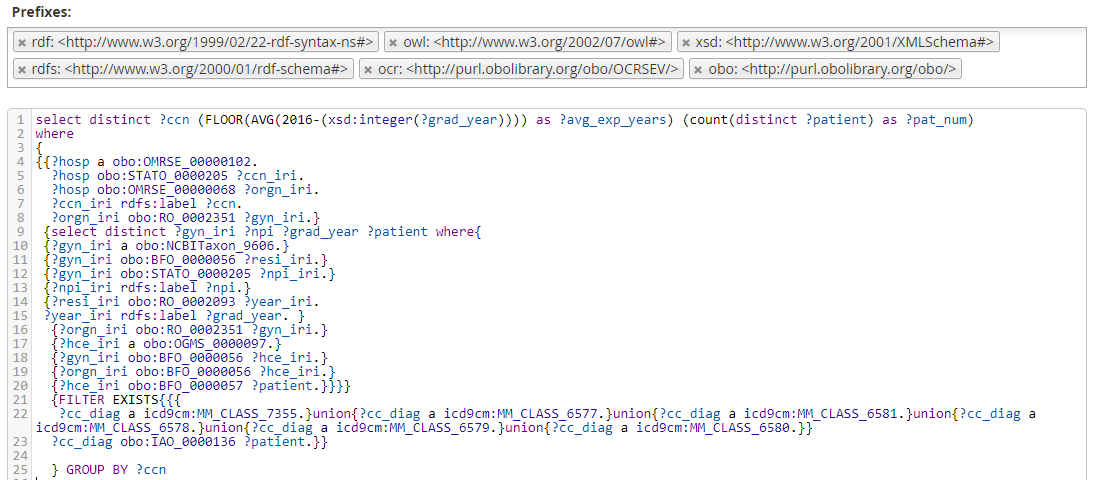
{?hce\_iri obo:BFO\_0000057 ?patient.}}}}

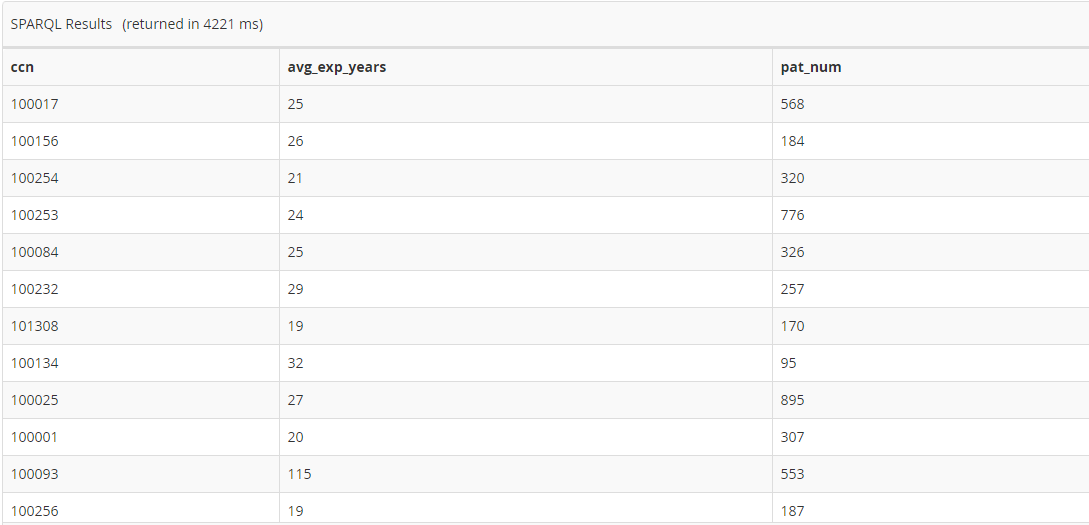
{FILTER EXISTS{{{

?cc\_diag a icd9cm:MM\_CLASS\_7355.}union{?cc\_diag a icd9cm:MM\_CLASS\_6577.}union{?cc\_diag a icd9cm:MM\_CLASS\_6581.}union{?cc\_diag a icd9cm:MM\_CLASS\_6578.}union{?cc\_diag a icd9cm:MM\_CLASS\_6579.}union{?cc\_diag a icd9cm:MM\_CLASS\_6580.}}

?cc\_diag obo:IAO\_0000136 ?patient.}}

} GROUP BY ?ccn





12. List the maximum recorded systolic and diastolic per CC patient.

select distinct ?patient (max(xsd:integer(strbefore(?bp, '/'))) as ?max\_systolic)

(max(xsd:integer(strafter(?bp, '/'))) as ?max\_diastolic)

where

{

{?bpmd a obo:VSO\_0000005.

?bpmd rdfs:label ?bp.

?bpmd obo:IAO\_0000136 ?bp\_iri.

?bp\_iri obo:RO\_0000052 ?patient.

}

{{{

?cc\_diag a icd9cm:MM\_CLASS\_7355.}union{?cc\_diag a icd9cm:MM\_CLASS\_6577.}union{?cc\_diag a icd9cm:MM\_CLASS\_6581.}union{?cc\_diag a icd9cm:MM\_CLASS\_6578.}union{?cc\_diag a icd9cm:MM\_CLASS\_6579.}union{?cc\_diag a icd9cm:MM\_CLASS\_6580.}}

?cc\_diag obo:IAO\_0000136 ?patient.}

FILTER(((xsd:integer(strbefore(?bp, '/'))) > 120)||((xsd:integer(strafter(?bp, '/'))) > 80))

} GROUP BY ?patient





13. List the CC patient who have elevated blood pressure (between 120/80 and 130/80).

select distinct ?patient ?bp

where

{

{?bpmd a obo:VSO\_0000005.

?bpmd rdfs:label ?bp.

?bpmd obo:IAO\_0000136 ?bp\_iri.

?bp\_iri obo:RO\_0000052 ?patient.

}

{{{

?cc\_diag a icd9cm:MM\_CLASS\_5861.}union{?cc\_diag a icd9cm:MM\_CLASS\_5563.}union{?cc\_diag a icd9cm:MM\_CLASS\_7355.}union{?cc\_diag a icd9cm:MM\_CLASS\_6577.}union{?cc\_diag a icd9cm:MM\_CLASS\_6581.}union{?cc\_diag a icd9cm:MM\_CLASS\_6578.}union{?cc\_diag a icd9cm:MM\_CLASS\_6579.}union{?cc\_diag a icd9cm:MM\_CLASS\_6580.}}

?cc\_diag obo:IAO\_0000136 ?patient.}

FILTER(((xsd:integer(strbefore(?bp, '/'))) > 120)&&((xsd:integer(strbefore(?bp, '/'))) < 130)&&((xsd:integer(strafter(?bp, '/'))) < 80))

}





14. Count the number of CC patients who have high blood pressure (> 130/80).

select (count(distinct ?patient) as ?pat\_num)

where

{

{?bpmd a obo:VSO\_0000005.

?bpmd rdfs:label ?bp.

?bpmd obo:IAO\_0000136 ?bp\_iri.

?bp\_iri obo:RO\_0000052 ?patient.

}

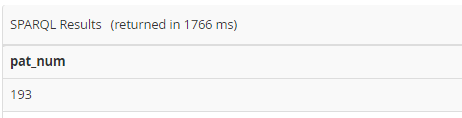
{{{ ?cc\_diag a icd9cm:MM\_CLASS\_7355.}union{?cc\_diag a icd9cm:MM\_CLASS\_6577.}union{?cc\_diag a icd9cm:MM\_CLASS\_6581.}union{?cc\_diag a icd9cm:MM\_CLASS\_6578.}union{?cc\_diag a icd9cm:MM\_CLASS\_6579.}union{?cc\_diag a icd9cm:MM\_CLASS\_6580.}}

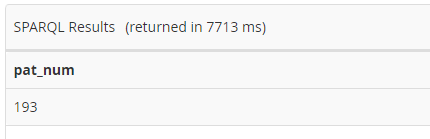
?cc\_diag obo:IAO\_0000136 ?patient.}

FILTER(((xsd:integer(strbefore(?bp, '/'))) > 129)||((xsd:integer(strafter(?bp, '/'))) > 79))

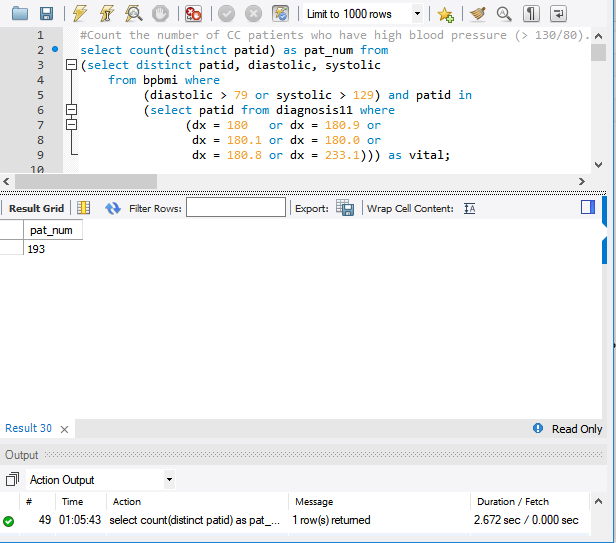
}







SQL Validation



15. Count the number of CC patients who are underweight.

select (count(distinct ?patient) as ?pat\_num)

where

{

{?bmi\_iri a obo:CMO\_0000105.

?bmi\_iri rdfs:label ?bmi.

?bmi\_iri obo:IAO\_0000136 ?bm\_iri.

?bm\_iri obo:RO\_0000052 ?patient.

}

{{{

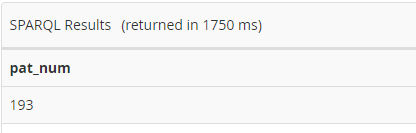
?cc\_diag a icd9cm:MM\_CLASS\_7355.}union{?cc\_diag a icd9cm:MM\_CLASS\_6577.}union{?cc\_diag a icd9cm:MM\_CLASS\_6581.}union{?cc\_diag a icd9cm:MM\_CLASS\_6578.}union{?cc\_diag a icd9cm:MM\_CLASS\_6579.}union{?cc\_diag a icd9cm:MM\_CLASS\_6580.}}

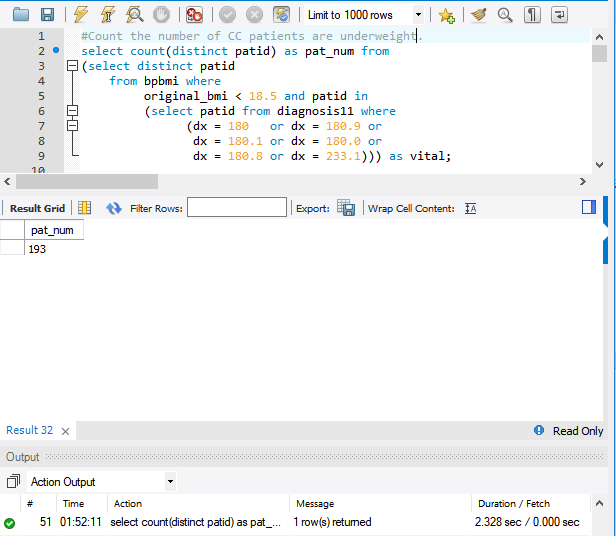
?cc\_diag obo:IAO\_0000136 ?patient.}

FILTER((xsd:float(?bmi)) < 18.5)

}







16. List the CC patients with their minimum and maximum recorded BMI.

select ?patient (max(xsd:float(?bmi)) as ?max\_bmi) (min(xsd:float(?bmi)) as ?min\_bmi)

where

{

{?bmi\_iri a obo:CMO\_0000105.

?bmi\_iri rdfs:label ?bmi.

?bmi\_iri obo:IAO\_0000136 ?bm\_iri.

?bm\_iri obo:RO\_0000052 ?patient.

}

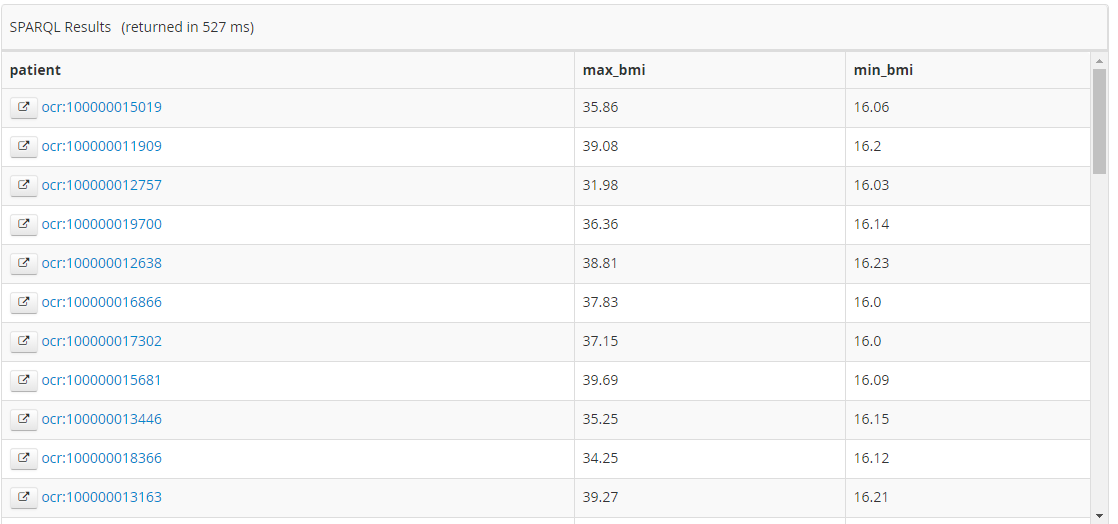
{{{

?cc\_diag a icd9cm:MM\_CLASS\_7355.}union{?cc\_diag a icd9cm:MM\_CLASS\_6577.}union{?cc\_diag a icd9cm:MM\_CLASS\_6581.}union{?cc\_diag a icd9cm:MM\_CLASS\_6578.}union{?cc\_diag a icd9cm:MM\_CLASS\_6579.}union{?cc\_diag a icd9cm:MM\_CLASS\_6580.}}

?cc\_diag obo:IAO\_0000136 ?patient.}

} GROUP BY ?patient





17. List the CC patients with abnormal BMI and high blood pressure along with BMI and BP taken during different encounters.

select distinct ?patient ?bmi ?bp

where

{

{?bmi\_iri a obo:CMO\_0000105.

?bmi\_iri rdfs:label ?bmi.

?bmi\_iri obo:IAO\_0000136 ?bm\_iri.

?bm\_iri obo:RO\_0000052 ?patient.

}

{?bpmd a obo:VSO\_0000005.

?bpmd rdfs:label ?bp.

?bpmd obo:IAO\_0000136 ?bp\_iri.

?bp\_iri obo:RO\_0000052 ?patient.

}

{{{

?cc\_diag a icd9cm:MM\_CLASS\_5861.}union{?cc\_diag a icd9cm:MM\_CLASS\_5563.}union{?cc\_diag a icd9cm:MM\_CLASS\_7355.}union{?cc\_diag a icd9cm:MM\_CLASS\_6577.}union{?cc\_diag a icd9cm:MM\_CLASS\_6581.}union{?cc\_diag a icd9cm:MM\_CLASS\_6578.}union{?cc\_diag a icd9cm:MM\_CLASS\_6579.}union{?cc\_diag a icd9cm:MM\_CLASS\_6580.}}

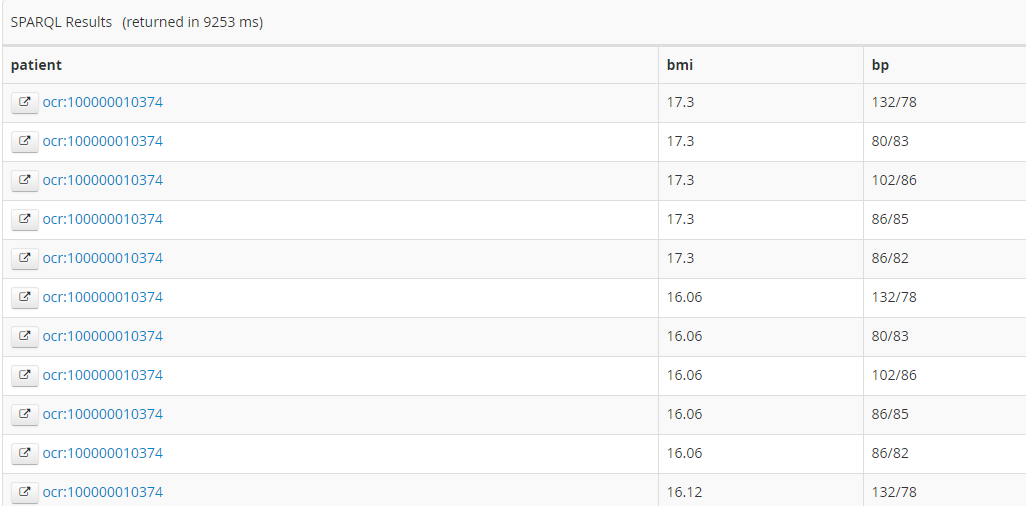
?cc\_diag obo:IAO\_0000136 ?patient.}

FILTER(((xsd:float(?bmi)) > 24.9) || ((xsd:float(?bmi)) < 18.5))

FILTER(((xsd:integer(strbefore(?bp, '/'))) > 129)||((xsd:integer(strafter(?bp, '/'))) > 79))

}





18. How many people got diagnosed with CC after the effective date of BCCPTA? List both the policies with their effective dates and the number of people who got diagnosed with CC after the effective dates.

select ?act ?policy\_date (count(distinct ?patient) as ?num\_pat)

where{

{

{

{ ?date\_iri a obo:IAO\_0000314.

?date\_iri obo:BFO\_0000050 ?doc\_iri.

?date\_iri obo:IAO\_0000136 ?temp\_iri.}

{?doc\_iri obo:OBI\_0000295 ?act\_iri.

?act\_iri rdfs:label ?act.}

{?temp\_iri rdfs:label ?policy\_date.}}

{{{ ?cc\_diag a icd9cm:MM\_CLASS\_7355.}union{?cc\_diag a icd9cm:MM\_CLASS\_6577.}union{?cc\_diag a icd9cm:MM\_CLASS\_6581.}union{?cc\_diag a icd9cm:MM\_CLASS\_6578.}union{?cc\_diag a icd9cm:MM\_CLASS\_6579.}union{?cc\_diag a icd9cm:MM\_CLASS\_6580.}}

{?cc\_diag obo:IAO\_0000136 ?patient.}

{?patient a obo:NCBITaxon\_9606.}

{?cc\_diag obo:OBI\_0000312 ?process\_iri.}

{?process\_iri obo:BFO\_0000155 ?intervl\_iri.}

{?intervl\_iri obo:RO\_0002091 ?dat\_iri.}

{?dat\_iri rdfs:label ?admit\_date.}}

}

FILTER(

(xsd:float(concat((strafter(strafter(?policy\_date, '/'), '/')),'.',(strbefore(?policy\_date, '/'))))) <

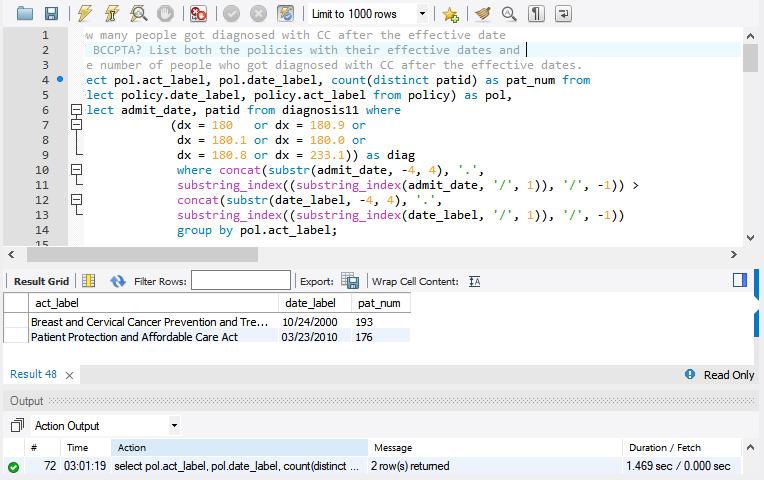
(xsd:float(concat((strafter(strafter(?admit\_date, '/'), '/')),'.',(strbefore(?admit\_date, '/'))))))

} GROUP BY ?act ?policy\_date





SQL Validation:



19. List all CC patients who also have a diagnosis of HIV having female biological sex and those who use tobacco along with their race, ethnicity, gender, sexual orientation, smoking prevalence in their MMSA region and the number of physicians they have had an encounter with.

select distinct ?patient ?race ?ethnicity ?gender\_value ?sexual\_ori\_value ?prev\_data (count(distinct ?gyn\_iri) as ?gyn\_num) where

{ {?race\_id a obo:OMRSE\_00000099.}{?race\_iri obo:OBI\_0000312 ?race\_id.}{?race\_iri a ?race.}{?race\_iri obo:IAO\_0000136 ?patient.}

{?eth\_id a obo:OMRSE\_00000101.}{?eth\_iri obo:OBI\_0000312 ?eth\_id.}{?eth\_iri a ?ethnicity.}{?eth\_iri obo:IAO\_0000136 ?patient.}

{?gen\_id a obo:OMRSE\_00000135.}{?gen\_iri obo:OBI\_0000312 ?gen\_id.}{?gen\_iri rdfs:label ?gender\_value.}{?gen\_iri obo:IAO\_0000136 ?patient.}

{?so\_id a obo:OMRSE\_00000142.}{?so\_iri obo:OBI\_0000312 ?so\_id.}{?so\_iri rdfs:label ?sexual\_ori\_value.}{?so\_iri obo:IAO\_0000136 ?patient.}

{?tu\_role a ocr:OCRSEV\_00000545.?tu\_role obo:RO\_0000052 ?patient.}

{?hiv\_diag a icd9cm:MM\_CLASS\_5563.?hiv\_diag obo:IAO\_0000136 ?patient.}

{{?smoke\_prev\_iri a ocr:OCRSEV\_00000583.?smoke\_prev\_iri rdfs:label ?prev\_data.}

{?smoke\_prev\_iri obo:IAO\_0000136 ?space.?space rdfs:label ?location.}

{?space obo:RO\_0001015 ?zip2d.?zip2d obo:RO\_0001015 ?patient.}}

{{?orgn\_iri obo:RO\_0002351 ?gyn\_iri.}{?gyn\_iri a obo:NCBITaxon\_9606.}{?gyn\_iri obo:BFO\_0000056 ?resi\_iri.}{?gyn\_iri obo:STATO\_0000205 ?npi\_iri.}

{?npi\_iri rdfs:label ?npi.}

{?resi\_iri obo:RO\_0002093 ?year\_iri.?year\_iri rdfs:label ?grad\_year. }

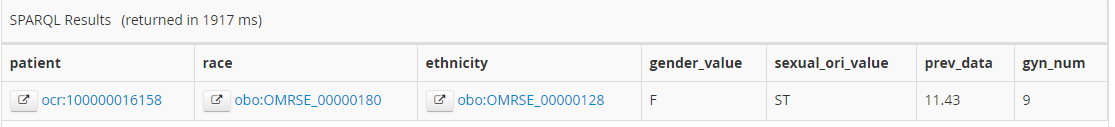
{?orgn\_iri obo:RO\_0002351 ?gyn\_iri.}

{?hce\_iri a obo:OGMS\_0000097.}{?gyn\_iri obo:BFO\_0000056 ?hce\_iri.}{?orgn\_iri obo:BFO\_0000056 ?hce\_iri.}{?hce\_iri obo:BFO\_0000057 ?patient.}}

{{{?cc\_diag a icd9cm:MM\_CLASS\_7355.}union{?cc\_diag a icd9cm:MM\_CLASS\_6577.}union{?cc\_diag a icd9cm:MM\_CLASS\_6581.}union{?cc\_diag a icd9cm:MM\_CLASS\_6578.}union{?cc\_diag a icd9cm:MM\_CLASS\_6579.}union{?cc\_diag a icd9cm:MM\_CLASS\_6580.}} ?cc\_diag obo:IAO\_0000136 ?patient.}

FILTER EXISTS { ?sex a obo:PATO\_0000383. ?sex obo:RO\_0000052 ?patient.}} GROUP BY ?patient ?race ?ethnicity ?gender\_value ?sexual\_ori\_value ?prev\_data





20. List all people who have HIV and who got diagnosed with CC after the effective date of BCCPTA along with their race, gender, sexual orientation, smoking prevalence in their MMSA region, the ccn of the hospitals they have had an encounter at, the MSPB score of those hospitals and the number of physicians per hospital they have had an encounter with.

This is the most complex query covering all five levels of SEM and it takes 6 minutes to finish.

select distinct ?ccn ?mspb\_score ?patient ?race ?gender\_value ?sexual\_ori\_value ?prev\_data (count(distinct ?gyn\_iri) as ?gyn\_num) where{{{?race\_id a obo:OMRSE\_00000099.}{?race\_iri obo:OBI\_0000312 ?race\_id.}{?race\_iri a ?race.}{?race\_iri obo:IAO\_0000136 ?patient.}{?gen\_id a obo:OMRSE\_00000135.}{?gen\_iri obo:OBI\_0000312 ?gen\_id.}{?gen\_iri rdfs:label ?gender\_value.}{?gen\_iri obo:IAO\_0000136 ?patient.}{?so\_id a obo:OMRSE\_00000142.}{?so\_iri obo:OBI\_0000312 ?so\_id.}{?so\_iri rdfs:label ?sexual\_ori\_value.}{?so\_iri obo:IAO\_0000136 ?patient.}

{?hiv\_diag a icd9cm:MM\_CLASS\_5563.?hiv\_diag obo:IAO\_0000136 ?patient.}

{{?smoke\_prev\_iri a ocr:OCRSEV\_00000583.?smoke\_prev\_iri rdfs:label ?prev\_data.}{?smoke\_prev\_iri obo:IAO\_0000136 ?space.?space rdfs:label ?location.}{?space obo:RO\_0001015 ?zip2d.?zip2d obo:RO\_0001015 ?patient.}}

{{?orgn\_iri obo:RO\_0002351 ?gyn\_iri.}{?gyn\_iri a obo:NCBITaxon\_9606.}{?gyn\_iri obo:BFO\_0000056 ?resi\_iri.}{?gyn\_iri obo:STATO\_0000205 ?npi\_iri.}{?npi\_iri rdfs:label ?npi.}{?resi\_iri obo:RO\_0002093 ?year\_iri.?year\_iri rdfs:label ?grad\_year. }{?orgn\_iri obo:RO\_0002351 ?gyn\_iri.}{?hce\_iri a obo:OGMS\_0000097.}{?gyn\_iri obo:BFO\_0000056 ?hce\_iri.}{?orgn\_iri obo:BFO\_0000056 ?hce\_iri.}{?hce\_iri obo:BFO\_0000057 ?patient.}}

{?hosp a obo:OMRSE\_00000102.?hosp obo:STATO\_0000205 ?ccn\_iri.?hosp obo:OMRSE\_00000068 ?orgn\_iri.?ccn\_iri rdfs:label ?ccn.?mspb a ocr:OCRSEV\_00000740.?mspb obo:IAO\_0000136 ?orgn\_iri.?mspb rdfs:label ?mspb\_score.{?gyn\_iri a obo:NCBITaxon\_9606.?orgn\_iri obo:RO\_0002351 ?gyn\_iri.}{?hce\_iri a obo:OGMS\_0000097.}{?gyn\_iri obo:BFO\_0000056 ?hce\_iri.}{?orgn\_iri obo:BFO\_0000056 ?hce\_iri.}{?hce\_iri obo:BFO\_0000057 ?patient.}}

{{{?cc\_diag a icd9cm:MM\_CLASS\_7355.}union{?cc\_diag a icd9cm:MM\_CLASS\_6577.}union{?cc\_diag a icd9cm:MM\_CLASS\_6581.}union{?cc\_diag a icd9cm:MM\_CLASS\_6578.}union{?cc\_diag a icd9cm:MM\_CLASS\_6579.}union{?cc\_diag a icd9cm:MM\_CLASS\_6580.}} ?cc\_diag obo:IAO\_0000136 ?patient.}{{?cc\_diag obo:IAO\_0000136 ?patient.}{?patient a obo:NCBITaxon\_9606.}{?cc\_diag obo:OBI\_0000312 ?process\_iri.}{?process\_iri obo:BFO\_0000155 ?intervl\_iri.}{?intervl\_iri obo:RO\_0002091 ?dat\_iri.}{?dat\_iri rdfs:label ?admit\_date.}}

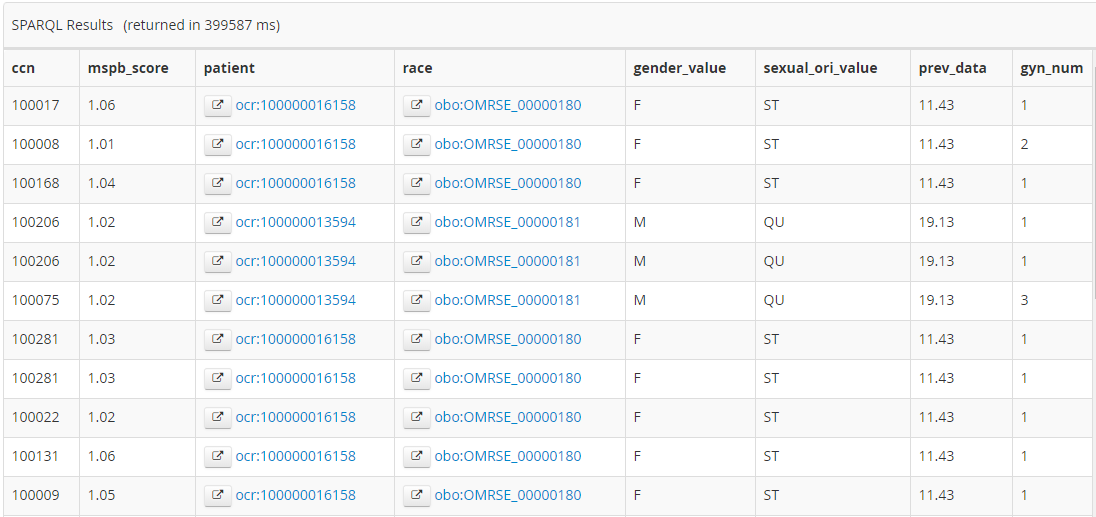
{{ ?date\_iri a obo:IAO\_0000314. ?date\_iri obo:BFO\_0000050 ?doc\_iri. ?date\_iri obo:IAO\_0000136 ?temp\_iri.}{?doc\_iri obo:OBI\_0000295 ?act\_iri. ?act\_iri rdfs:label 'Breast and Cervical Cancer Prevention and Treatment Act'.}{?temp\_iri rdfs:label ?policy\_date.}}

}FILTER((xsd:float(concat((strafter(strafter(?policy\_date, '/'), '/')),'.',(strbefore(?policy\_date, '/'))))) <

(xsd:float(concat((strafter(strafter(?admit\_date, '/'), '/')),'.',(strbefore(?admit\_date, '/'))))))

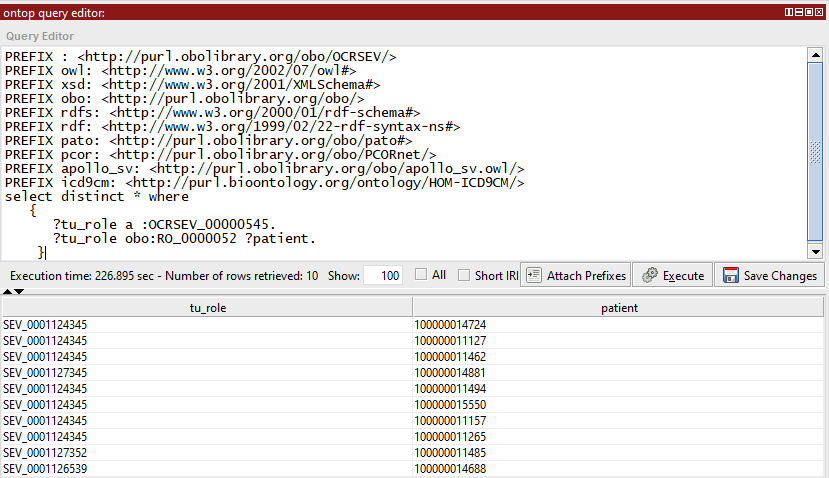
} GROUP BY ?ccn ?mspb\_score ?patient ?race ?gender\_value ?sexual\_ori\_value ?prev\_data



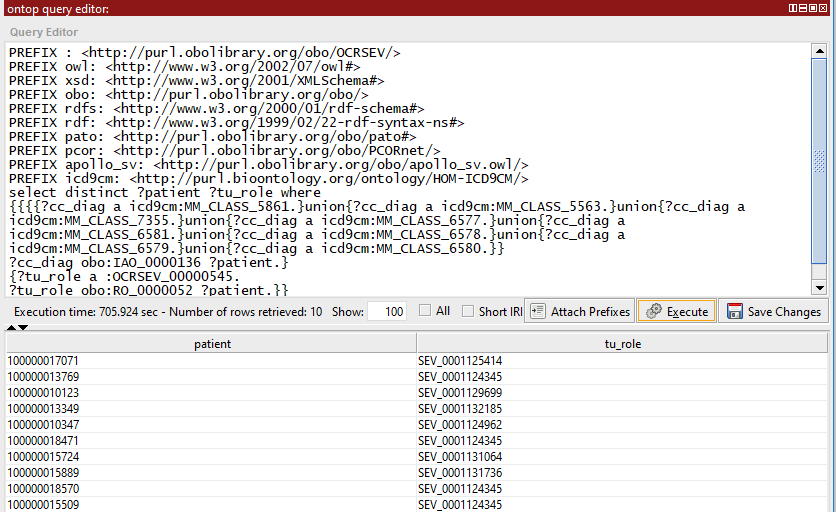


ONTOP SPARQL:

List all patients who are tobacco users:

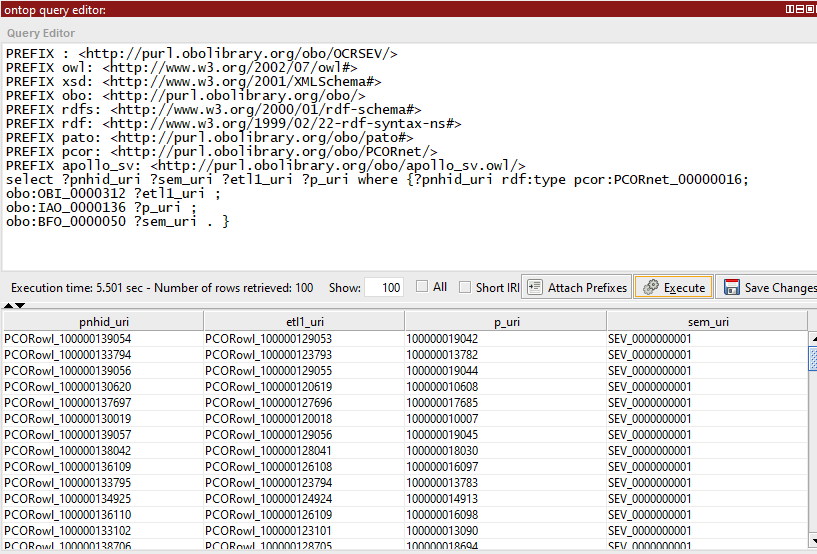


List all CC patients who are tobacco users: (11 minutes) – the same query runs only for 90 milli seconds in stardog.



3. Basic query constructed based one mapping: List all instances of PCORnet Hispanic identity datum along with the level of CRCCP SEM, its corresponding ETL instances and the patients associated with these instances.

This query took 29 minutes before RDB optimization. Now, it takes 5 seconds or less.



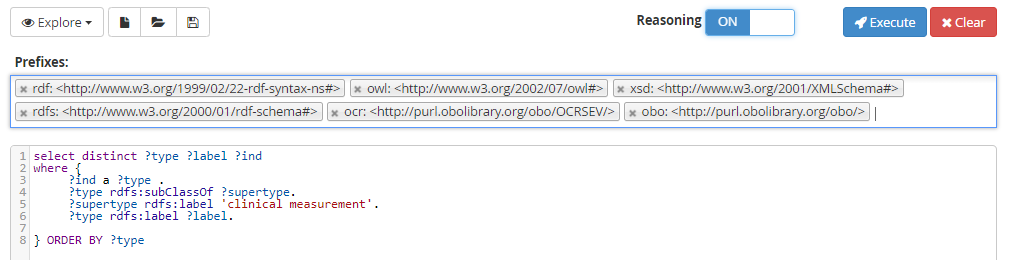
Queries to prove the effect of OCRSEV in semantic data integration:

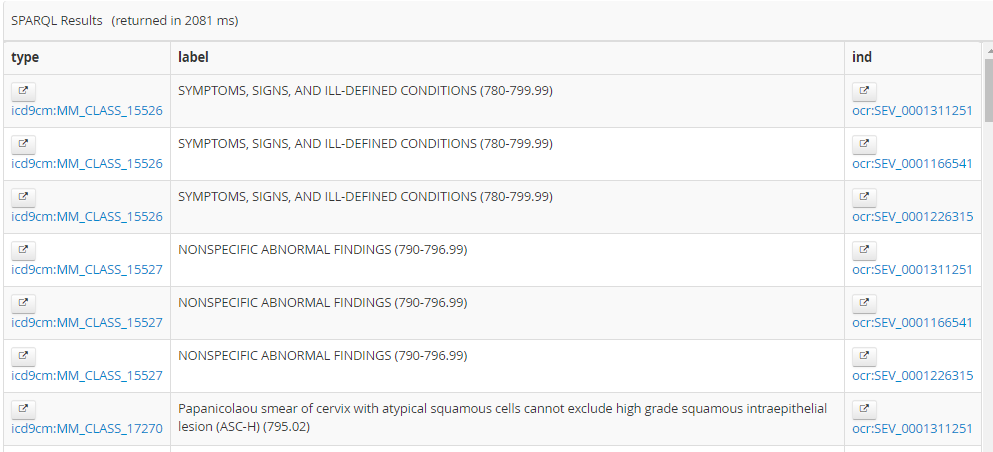
1. Query to show inheritance

Which instances are instances of subclasses of ‘clinical measurement’?

* The results should include all instances of BMI, diagnosis and its subclasses

|  |
| --- |
| select distinct ?type ?label ?ind |
|  | where { |
|  | ?ind a ?type . |
|  | ?type rdfs:subClassOf ?supertype . |
|  | ?supertype rdfs:label 'clinical measurement'. |
|  | ?type rdfs:label ?label. |
|  |  |
|  | } ORDER BY ?type |





2. Query to show DL expressivity in OCRSEV

Which class is defined by the “ ‘has participant’ some ‘Homo sapiens’ “ axiom and what are its instances?

According to the ontology models, “screening for malignant neoplasm of cervix” satisfies the condition mentioned in the query.

select ?type ?label ?ind

where {

?ind a ?type.

?type rdfs:subClassOf ?supertype.

?supertype a owl:Restriction ;

owl:onProperty obo:BFO\_0000057 ;

owl:someValuesFrom obo:NCBITaxon\_9606

.

?type rdfs:label ?label.

}



