The IT Operations Ontology (ITOPS) resource

R. Uceda-Sosa, N. Mihindukulasooriya, S. Bansal, S. Nagar, A. Kumar, V. Agarwal, Gaetano Rossiello, A. Gliozzo
IBM Research

Questions/comments: rosariou@us.ibm.com

Table of Contents

Introduction

We have built a domain-agnostic pipeline for domain-specific ontologies that leverages information on Wikidata, Wikipedia and DBPedia. Each of the three stages of the pipeline extends the previous ones by using a variety of symbolic and ML/DL techniques.

In this document we discuss the outcome of each stage and show sample SPARQL queries. We also discuss the content and topology of the GLO upper library of objects.

The product of each stage is a turtle (.ttl) file, ITOPS_S1.ttl, ITOPS_S2.ttl and ITOPS_S3.ttl, each extending the prior one. All of them build upon a General Library of Objects (GLO), which provides a general ontological framework to the domain specific graphs.

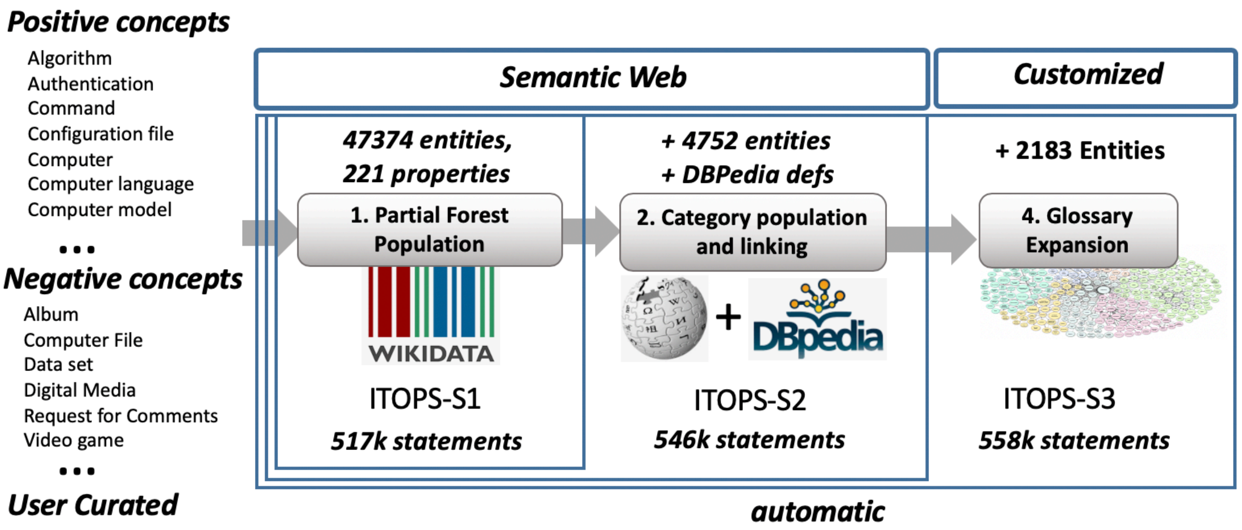


Figure 1. The pipeline used to build ITOPS

Each of the provided files can be installed in a triplestore and inferencing enginge that supports RDF/OWL (Blazegraph,Allegro,etc.)

CAVEAT -- It is worth noting that we've used a private installation of Wikidata using an RDF/OWL dump from late January, 2020, so there may be instances where our results differ from those in the live service.

The General Library of Objects

We have created a small, general T-Box describing domain entities (glo:DomainEntity), statements about these entities and a set of high-level relations among them. We have populated it with 74 general-purpose entities derived from Wikidata. These entities are instances of glo:DomainEntity and are mostly related by a partial order relation, glo:subConceptOf, which subsumes wdt:P279 (subclass of) and wdt:P31 (instance of). This is done so that subsequent extensions to the ontology don't need to redraw the boundaries between T-Box and A-Box.

We call this T-Box plus the seed A-Box the General Library of Objects (GLO). A snapshot is depicted below.

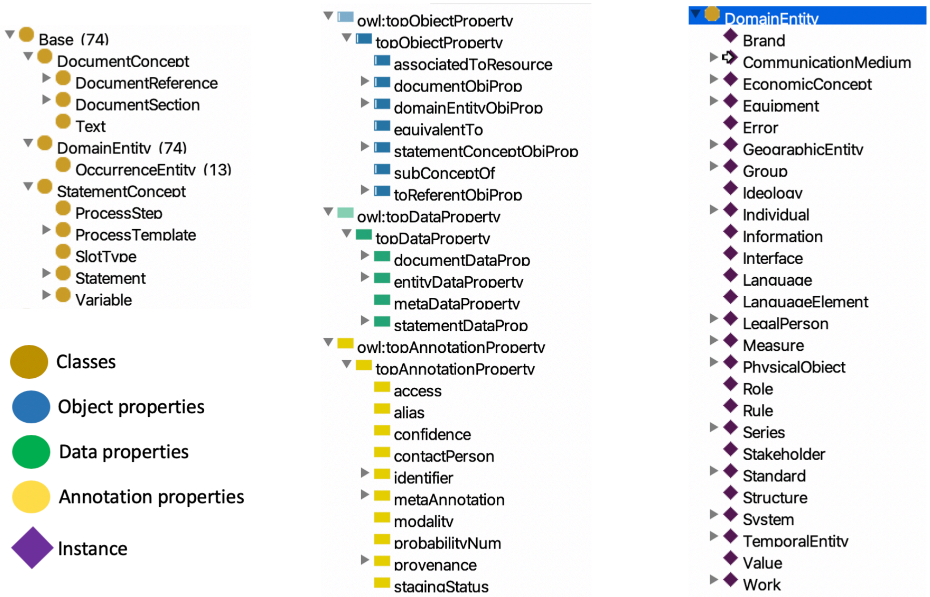
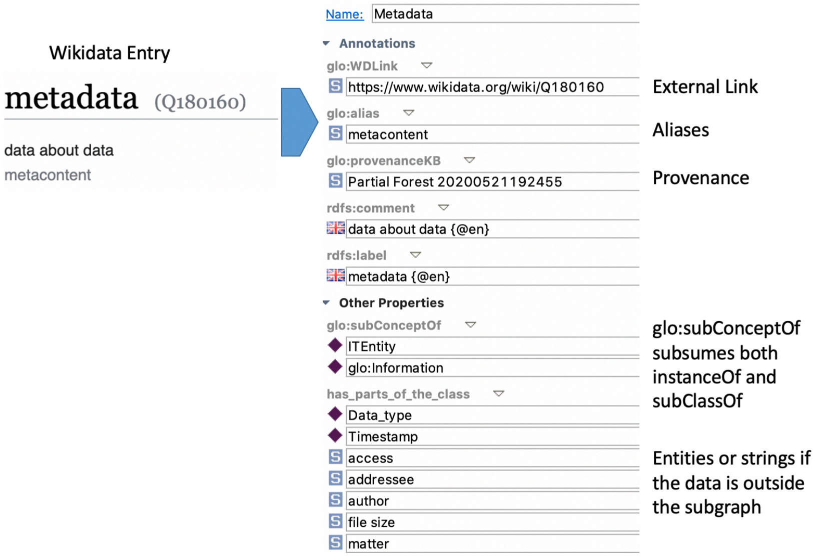


Figure 2. GLO Upper Ontology

The GLO a shortcut for http://www.ibm.com/GLO# . In order to understand GLO's content better, let's look at a sample entity.



Name: Metadata

Annotations

glo:WDLink
https://www.wikidata.org/wiki/Q180160

glo:alias
metacontent

glo:provenanceKB
Partial Forest 20200521192455

rdfs:comment
data about data {@en}

rdfs:label
metadata {@en}

Other Properties

glo:subConceptOf
ITEntity
glo:Information

has_parts_of_the_class
Data_type
Timestamp
access
addressee
author
file size
matter

Wikidata Entry
metadata (Q180160)
data about data
metacontent
External Link
Aliases
Provenance
glo:subConceptOf subsumes both instanceOf and subClassOf
Entities or strings if the data is outside the subgraph

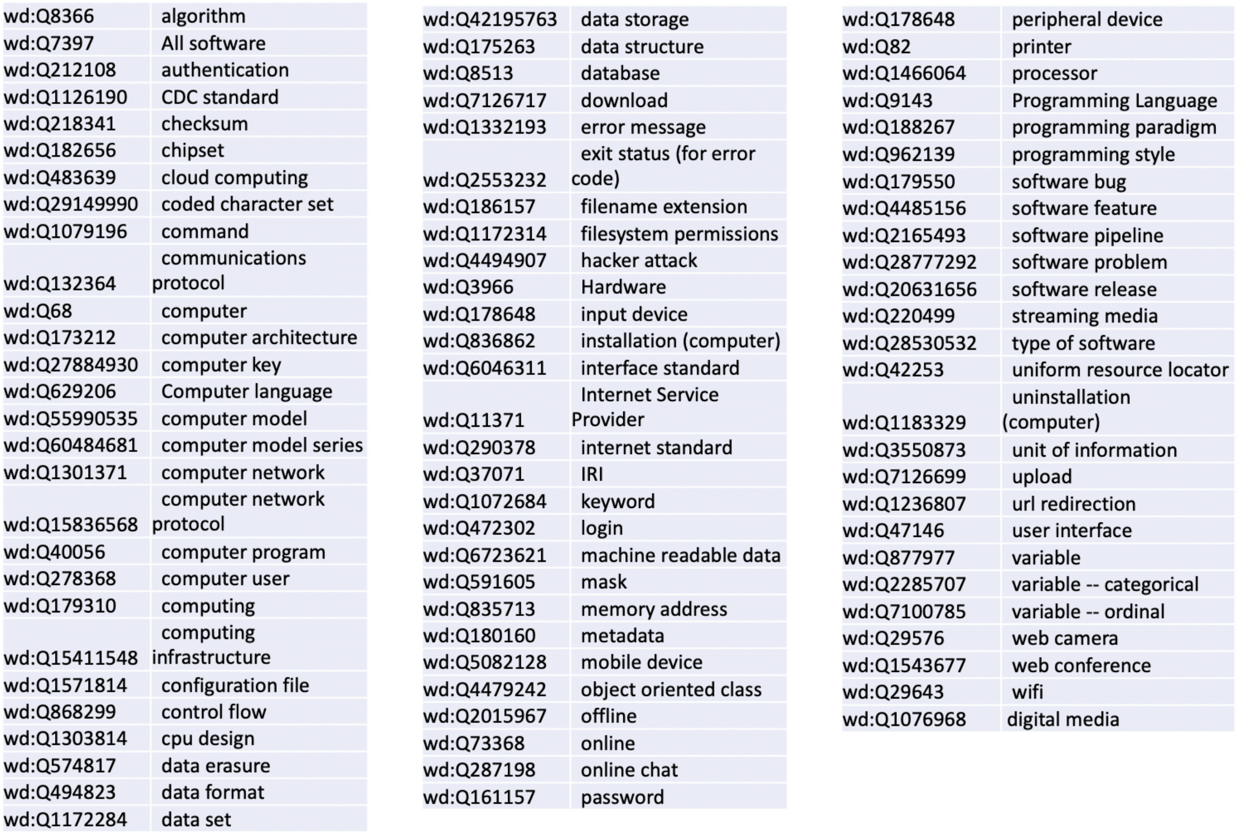
Figure 3. Sample entity using GLO vocabulary

Let's consider the entity "Metadata" in ITOPS, which uses the GLO. Each entity has metadata annotation properties associated to it, like aliases, comments, provenance (in this case indicating the step used to produce the entity) and external links (to Wikidata in this case). The relation glo:subConceptOf indicates that itops:Metadata is both an ITEntity and glo:Information. We discuss below how the inheritance hierarchy is constructed.

The relation itops:has_parts_of_the_class is one of the relations induced from the domain. These are defined as rdfs:subPropertyOf glo:domainEntityObjProp. All relations in a GLO ontology are categorized depending on whether they related domain entities (glo:domainEntityObjProp), statements (glo:statementConceptObjProp) or link statements to the domain entities they reference (glo:toReferentObjProp). Notice that some of the objects are instances, while others are strings. This is due to the construction pipeline, which seeks to limit the subgraph to relevant entities and avoids dangling references by referring to entities outside the scope of the ontology by strings.

S1 – The file ITOPS_S1.ttl

This file defines the itops namespace (http://www.ibm.com/ITOPS#) and is generated from a list of 82 common concepts in the TechQA dataset (https://arxiv.org/abs/1911.02984).

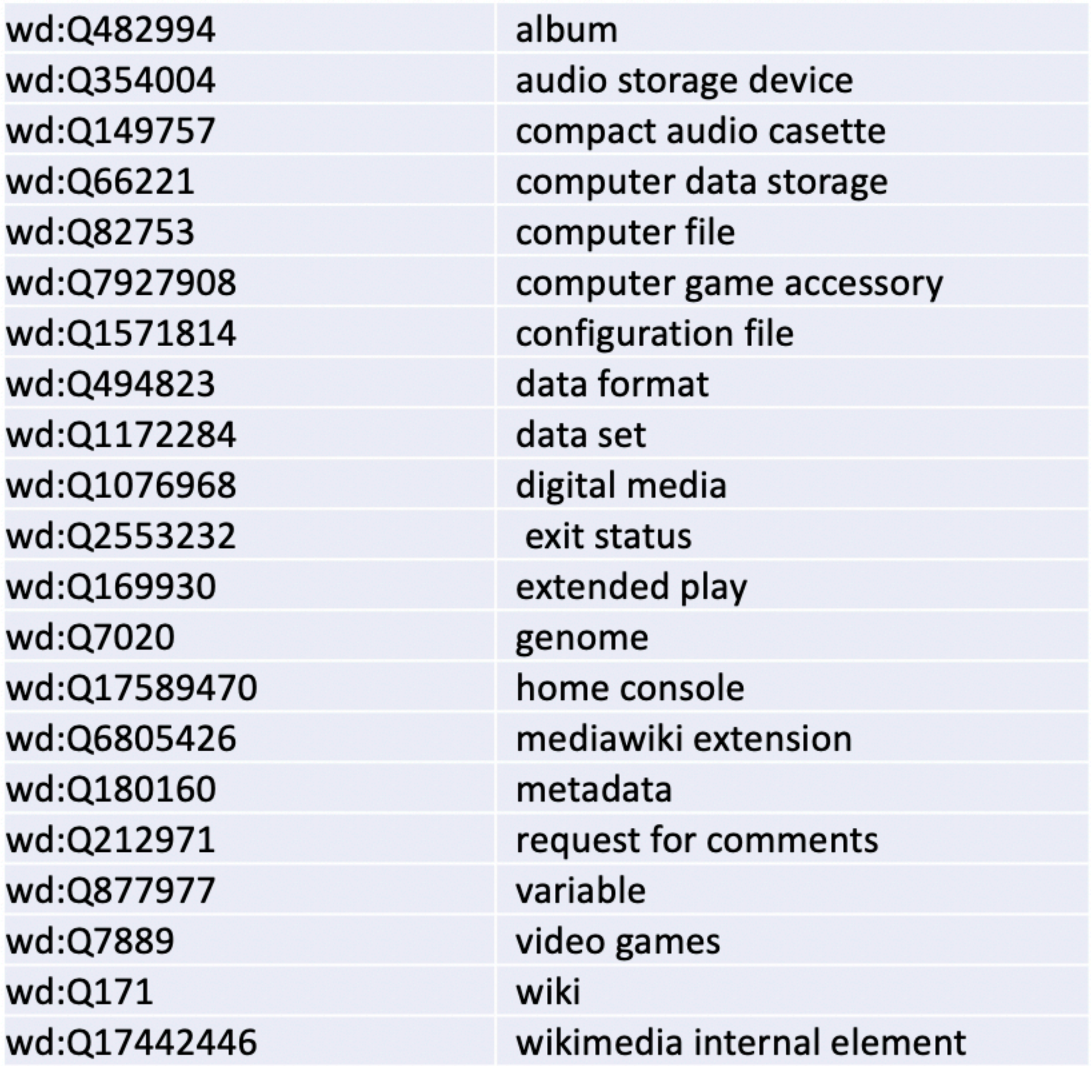


wd:Q8366	algorithm
wd:Q7397	All software
wd:Q212108	authentication
wd:Q1126190	CDC standard
wd:Q218341	checksum
wd:Q182656	chipset
wd:Q483639	cloud computing
wd:Q29149990	coded character set
wd:Q1079196	command
wd:Q132364	communications protocol
wd:Q68	computer
wd:Q173212	computer architecture
wd:Q27884930	computer key
wd:Q629206	Computer language
wd:Q55990535	computer model
wd:Q60484681	computer model series
wd:Q1301371	computer network
wd:Q15836568	computer network protocol
wd:Q40056	computer program
wd:Q278368	computer user
wd:Q179310	computing
wd:Q15411548	computing infrastructure
wd:Q1571814	configuration file
wd:Q868299	control flow
wd:Q1303814	cpu design
wd:Q574817	data erasure
wd:Q494823	data format
wd:Q1172284	data set
wd:Q42195763	data storage
wd:Q175263	data structure
wd:Q8513	database
wd:Q7126717	download
wd:Q1332193	error message
wd:Q2553232	exit status (for error code)
wd:Q186157	filename extension
wd:Q1172314	filesystem permissions
wd:Q4494907	hacker attack
wd:Q3966	Hardware
wd:Q178648	input device
wd:Q836862	installation (computer)
wd:Q6046311	interface standard
wd:Q11371	Internet Service Provider
wd:Q290378	internet standard
wd:Q37071	IRI
wd:Q1072684	keyword
wd:Q472302	login
wd:Q6723621	machine readable data
wd:Q591605	mask
wd:Q835713	memory address
wd:Q180160	metadata
wd:Q5082128	mobile device
wd:Q4479242	object oriented class
wd:Q2015967	offline
wd:Q73368	online
wd:Q287198	online chat
wd:Q161157	password
wd:Q178648	peripheral device
wd:Q82	printer
wd:Q1466064	processor
wd:Q9143	Programming Language
wd:Q188267	programming paradigm
wd:Q962139	programming style
wd:Q179550	software bug
wd:Q4485156	software feature
wd:Q2165493	software pipeline
wd:Q28777292	software problem
wd:Q20631656	software release
wd:Q220499	streaming media
wd:Q28530532	type of software
wd:Q42253	uniform resource locator
wd:Q1183329	uninstallation (computer)
wd:Q3550873	unit of information
wd:Q7126699	upload
wd:Q1236807	url redirection
wd:Q47146	user interface
wd:Q877977	variable
wd:Q2285707	variable -- categorical
wd:Q7100785	variable -- ordinal
wd:Q29576	web camera
wd:Q1543677	web conference
wd:Q29643	wifi
wd:Q1076968	digital media

Figure 4. Seed concepts for S1

From these seed terms we generate the relevant subgraph from Wikidata and brought as children of itops:. The top level entities are then 'linked' to the GLO ontology by following the wdt:P31 and wdt:P279 relations. In Figure 3, itops:Metadata is a subConcept of both itops:ITEntity and glo:Information. Notice that the intermediate concepts (those between the top layer of ITOPS and the lower layers of GLO) are ignored at this time.

It is worth noting that a population from these terms generates a large ontology with many useless hierarchies. In particular, music albums (https://www.wikidata.org/wiki/Q482994), videogames (https://www.wikidata.org/wiki/Q7889) or datasets (https://www.wikidata.org/wiki/Q117284) introduce thousands of instances with little use in the IT operations domain. After manually pruning some of the sections with an editor, we've compiled a list (DoNotPopulate) so subsequent runs do not include these instances. In the case of S1 the list is:



wd:Q482994	album
wd:Q354004	audio storage device
wd:Q149757	compact audio casette
wd:Q66221	computer data storage
wd:Q82753	computer file
wd:Q7927908	computer game accessory
wd:Q1571814	configuration file
wd:Q494823	data format
wd:Q1172284	data set
wd:Q1076968	digital media
wd:Q2553232	exit status
wd:Q169930	extended play
wd:Q7020	genome
wd:Q17589470	home console
wd:Q6805426	mediawiki extension
wd:Q180160	metadata
wd:Q212971	request for comments
wd:Q877977	variable
wd:Q7889	video games
wd:Q171	wiki
wd:Q17442446	wikimedia internal element

Fig 5. DoNotPopulate() list

Because of this DoNotPopulate list, we call this process "Partial Forest" population. In the glo:provenanceKB of the items obtained like this, we include this term and a timestamp.

The outcome of this step is a file with 47374 entities and 220 domain entity relation (plus subConceptOf). In the last section we show a few sample queries to obtain topology information at the end of this document.

S2 – The file ITOPS_S2.ttl

The S2 version of ITOPS seeks to leverage the categories from the English Wikipedia. Most Wikipedia articles are manually labelled with tags corresponding to categories. Such tags can contain information that is not explicitly in Wikidata. Most entities in Wikidata can then be linked to their corresponding Wikipedia article through interwiki information. This allows us to get the associated Wikipedia categories for each Wikidata entity.

For instance, if we take an entity such as “computer keyboard” (Q250), the corresponding Wikipedia page is tagged with four categories “computer peripherals”, “computing input devices”, “flexible electronics”, and “game control methods”. While the first two are domain-specific categories containing mostly IT domain entities, the latter two are more generic. Thus, we need a method for distinguishing between those.

We have used metrics that indicate the heterogeneity of entities in a category and their overlap with the initial partial forest to identify domain-specific cate- gories. For that, we extract all Wikipedia articles tagged with a given category and collect their corresponding entities along with their class types. For exam- ple, the entities “computer peripherals” are typed as instances of few classes that subclasses of “peripheral equipment” (Q178648) class while entities in “flexible electronics” are typed from many diverse classes that are far from each other in the class hierarchy. Furthermore, a large portion (75%) of entities in ‘computer peripherals’ overlap with the initial partial forest while for ‘flexible electronics’ it is around 9%.

The result is a file with 51096 entities and 220 domain entity relation (plus subConceptOf, of course). In the last section we show a few sample queries to obtain topology information at the end of this document.

S3 – The file ITOPS_S3.ttl

S2 does not take care of customization of ITOPS ontology with domain specific resources. IT domain has glossaries readily available for multiple sub-domains. These glossaries capture specialized knowledge of a domain and are usually a good quality data source.

S3 uses Deep Learning to classify concepts from glossaries along the glo:subConceptOf hierarchy in the ontology.

Four such glossaries are used:

1. https://www.ibm.com/support/knowledgecenter/STXNRM_3.14.7/coss.doc/
2. https://www.ibm.com/support/knowledgecenter/SSEPGG_11.1.0/com.ibm.db2.luw.glossary.doc/doc/glossary.html
3. https://www.dpsolutions.com/success-center/it-terminology-glossary
4. https://flexsystem.lenovofiles.com/help

Due to IP concerns, only the name and external link to the terms are stored in ITOPS_S3.

The result is a file with 53193 entities and 220 domain entity relation (plus subConceptOf, of course). In the last section we show a few sample queries to obtain topology information at the end of this document.

How to query ITOPS

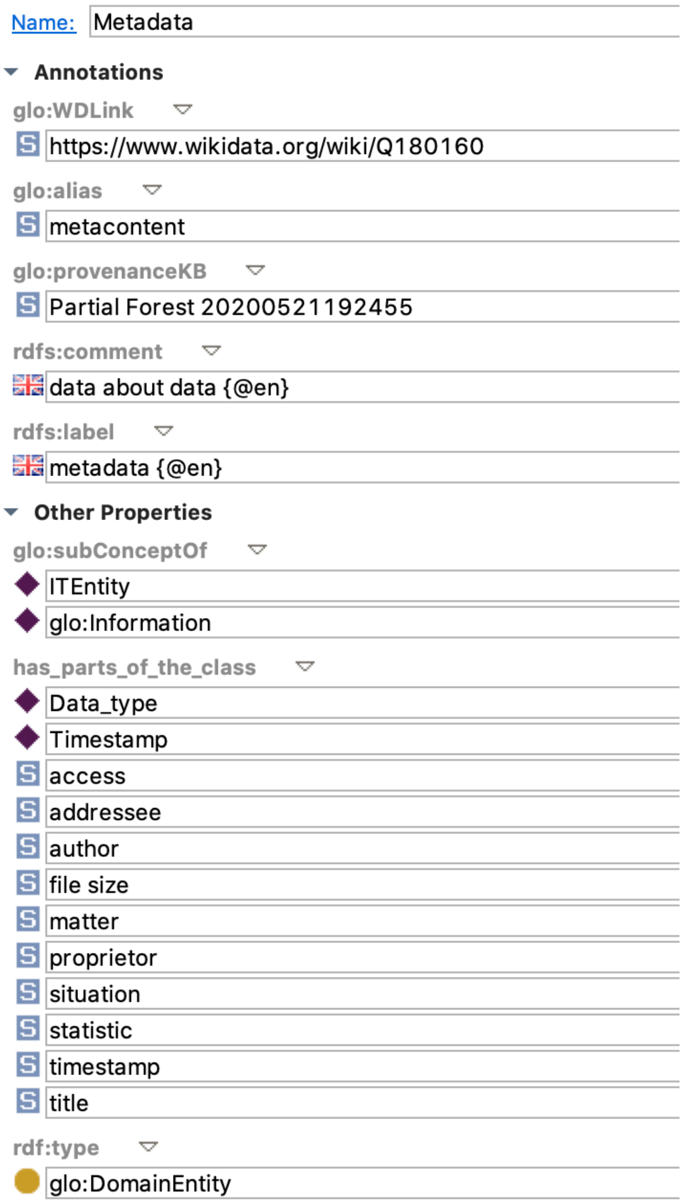
These queries can be run either of the three files, as the three stages result in an ontology with the <http://www.ibm.com/ITOPS# namespace. The examples provided here are from S1

QUERY: To obtain all properties in the graph (object, annotation, datatype)

```
prefix glo: <http://www.ibm.com/GLO#>
prefix itops: <http://www.ibm.com/ITOPS#>
SELECT DISTINCT ?property ?propertyLabel WHERE {
        {
                ?property rdfs:subPropertyOf glo:topObjectProperty .
         } UNION
        {
                 ?property rdfs:subPropertyOf glo:topDatatypeProperty .
        }
         UNION
        {
                 ?property rdfs:subPropertyOf u2o:topAnnotationProperty .
         }
```

QUERY: To obtain data for a given entity – outgoing properties

We select only the children properties and entities



Name: Metadata

Annotations

glo:WDLink
https://www.wikidata.org/wiki/Q180160

glo:alias
metacontent

glo:provenanceKB
Partial Forest 20200521192455

rdfs:comment
data about data {@en}

rdfs:label
metadata {@en}

Other Properties

glo:subConceptOf
ITEntity
glo:Information

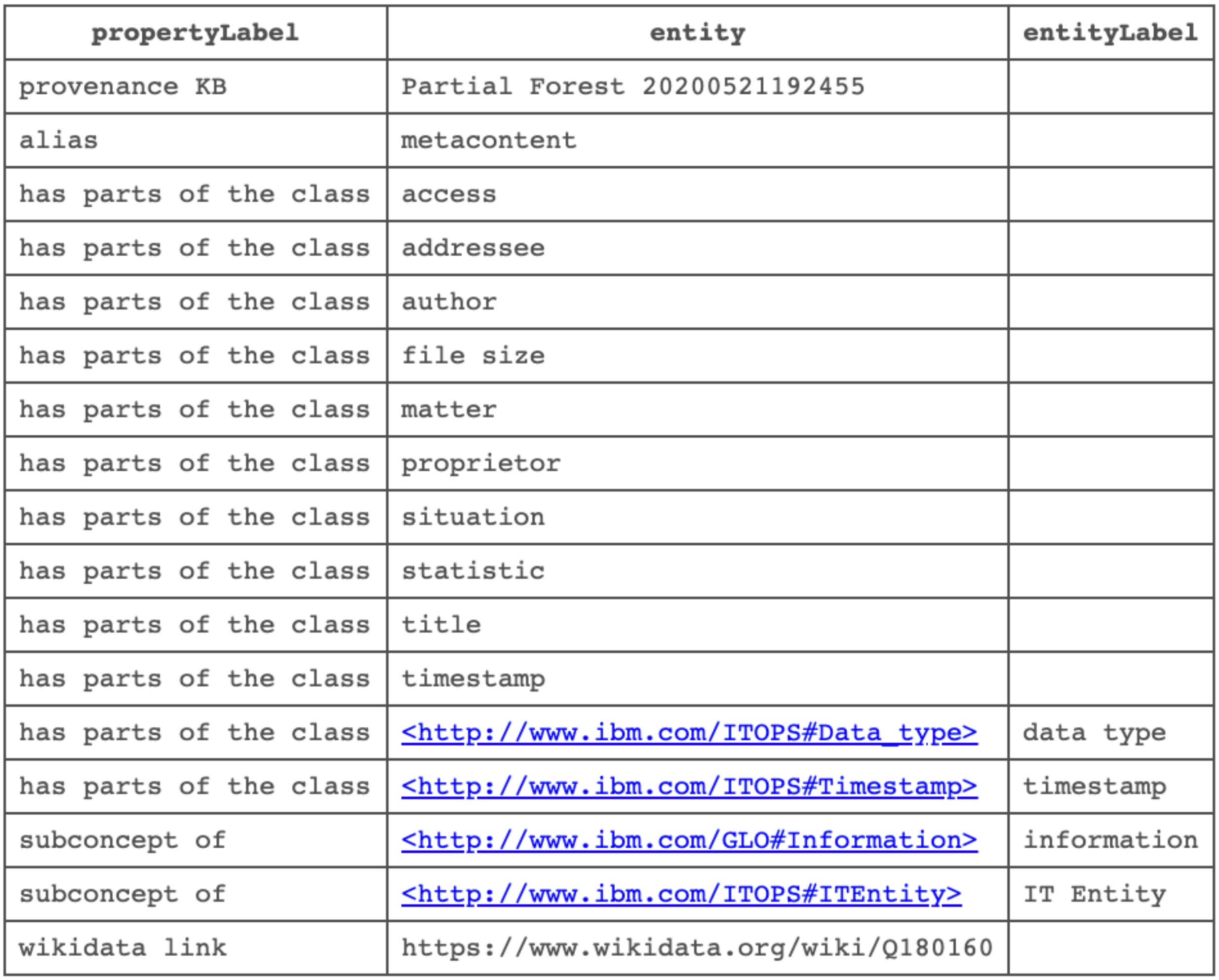
has_parts_of_the_class
Data_type
Timestamp
access
addressee
author
file size
matter
proprietor
situation
statistic
timestamp
title

rdf:type
glo:DomainEntity

```
prefix itops: <http://www.ibm.com/ITOPS#>
prefix glo: <http://www.ibm.com/GLO#>
SELECT DISTINCT ?propertyLabel ?entity ?entityLabel WHERE {
        itops:Metadata ?property ?entity .
        ?property rdfs:label ?propertyLabel .
         OPTIONAL {
                ?entity rdfs:label ?entityLabel .
        }
        # The most specific entity
        FILTER NOT EXISTS { ?otherEntity glo:subConceptOf ?entity .
             itops:Metadata ?property ?otherEntity .
```

```
                FILTER (?otherEntity != ?entity)
            }
  # The most specific property
  FILTER NOT EXISTS {?otherProperty rdfs:subPropertyOf ?property.
                itops:Metadata ?otherProperty ?entity .
                FILTER (?otherProperty != ?property)
               }
} ORDER BY ?propertyLabel
```

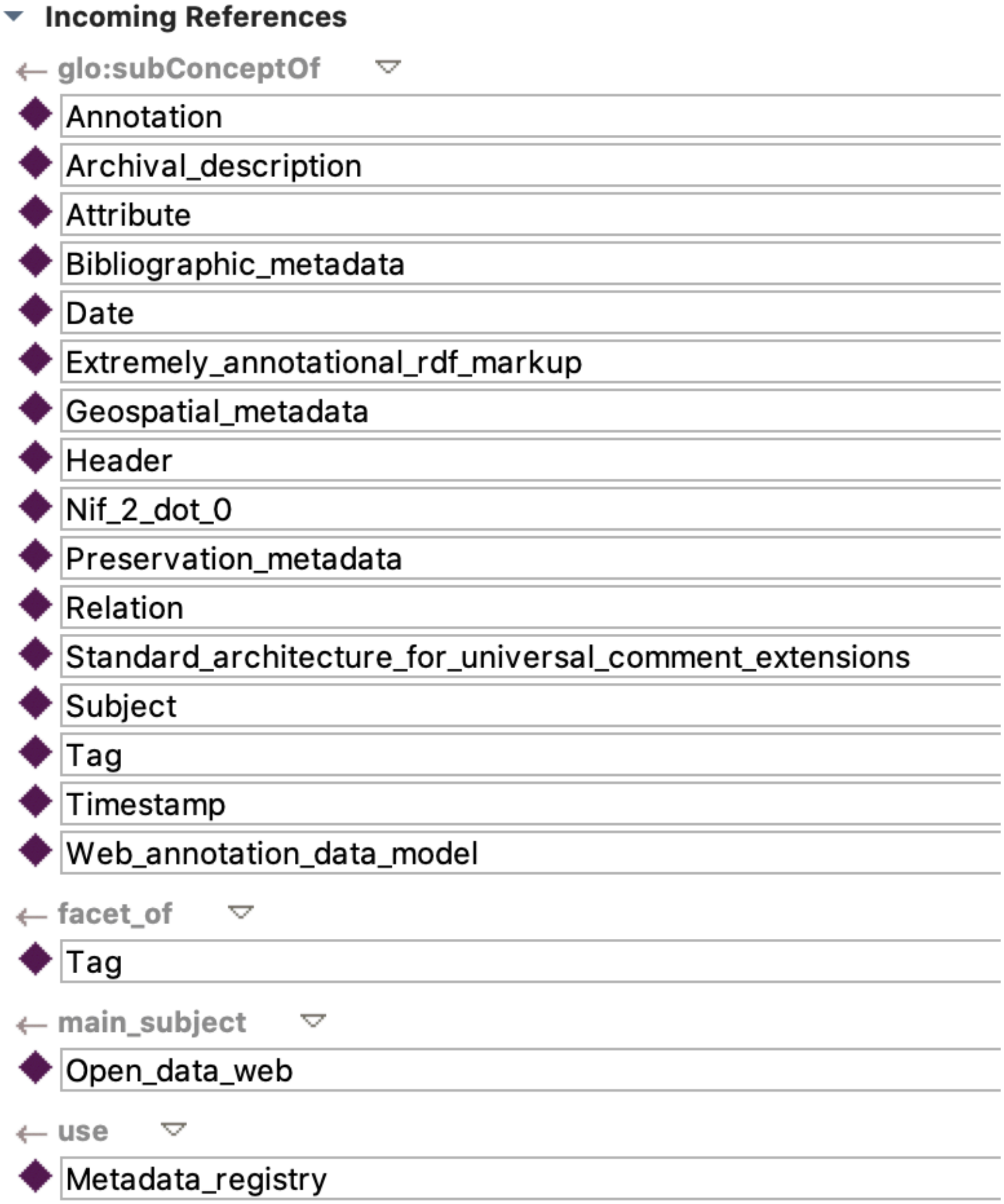
The results:



propertyLabel	entity	entityLabel
provenance KB	Partial Forest 20200521192455	
alias	metacontent	
has parts of the class	access	
has parts of the class	addressee	
has parts of the class	author	
has parts of the class	file size	
has parts of the class	matter	
has parts of the class	proprietor	
has parts of the class	situation	
has parts of the class	statistic	
has parts of the class	title	
has parts of the class	timestamp	
has parts of the class	<http://www.ibm.com/ITOPS#Data_type>	data type
has parts of the class	<http://www.ibm.com/ITOPS#Timestamp>	timestamp
subconcept of	<http://www.ibm.com/GLO#Information>	information
subconcept of	<http://www.ibm.com/ITOPS#ITEntity>	IT Entity
wikidata link	https://www.wikidata.org/wiki/Q180160	

QUERY: To obtain data for a given entity – incoming references

Using the same example from itops:Metadata



Incoming References

← glo:subConceptOf

- Annotation
- Archival_description
- Attribute
- Bibliographic_metadata
- Date
- Extremely_annotational_rdf_markup
- Geospatial_metadata
- Header
- Nif_2_dot_0
- Preservation_metadata
- Relation
- Standard_architecture_for_universal_comment_extensions
- Subject
- Tag
- Timestamp
- Web_annotation_data_model

← facet_of

- Tag

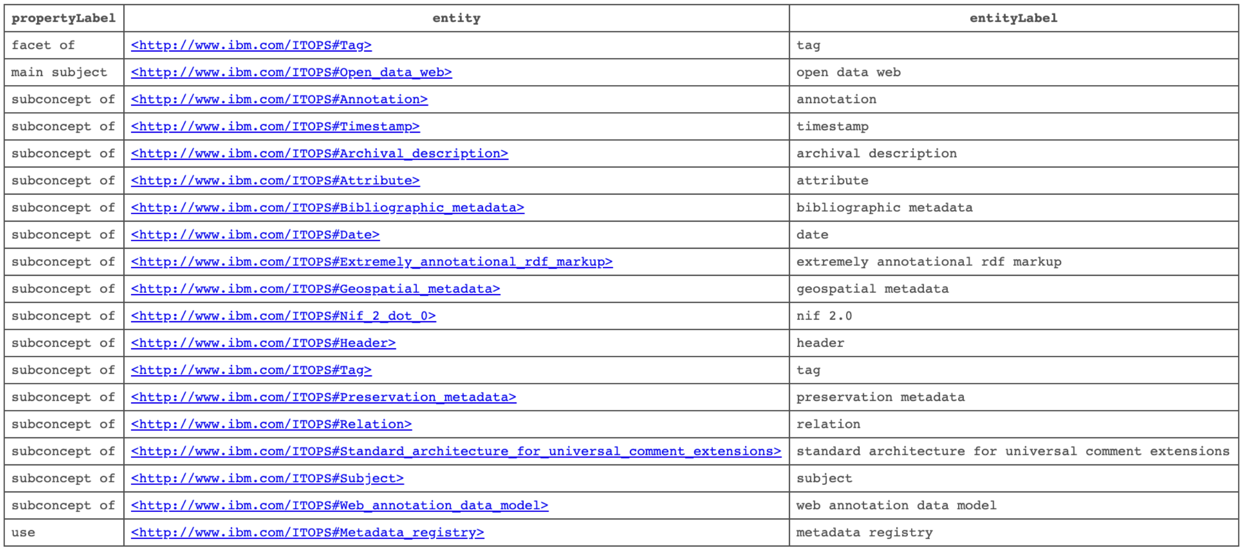
← main_subject

- Open_data_web

← use

- Metadata_registry

The result:



propertyLabel	entity	entityLabel
facet of	<http://www.ibm.com/ITOPS#Tag>	tag
main subject	<http://www.ibm.com/ITOPS#Open_data_web>	open data web
subconcept of	<http://www.ibm.com/ITOPS#Annotation>	annotation
subconcept of	<http://www.ibm.com/ITOPS#Timestamp>	timestamp
subconcept of	<http://www.ibm.com/ITOPS#Archival_description>	archival description
subconcept of	<http://www.ibm.com/ITOPS#Attribute>	attribute
subconcept of	<http://www.ibm.com/ITOPS#Bibliographic_metadata>	bibliographic metadata
subconcept of	<http://www.ibm.com/ITOPS#Date>	date
subconcept of	<http://www.ibm.com/ITOPS#Extremely_annotational_rdf_markup>	extremely annotational rdf markup
subconcept of	<http://www.ibm.com/ITOPS#Geospatial_metadata>	geospatial metadata
subconcept of	<http://www.ibm.com/ITOPS#Nif_2_dot_0>	nif 2.0
subconcept of	<http://www.ibm.com/ITOPS#Header>	header
subconcept of	<http://www.ibm.com/ITOPS#Tag>	tag
subconcept of	<http://www.ibm.com/ITOPS#Preservation_metadata>	preservation metadata
subconcept of	<http://www.ibm.com/ITOPS#Relation>	relation
subconcept of	<http://www.ibm.com/ITOPS#Standard_architecture_for_universal_comment_extensions>	standard architecture for universal comment extensions
subconcept of	<http://www.ibm.com/ITOPS#Subject>	subject
subconcept of	<http://www.ibm.com/ITOPS#Web_annotation_data_model>	web annotation data model
use	<http://www.ibm.com/ITOPS#Metadata_registry>	metadata registry

QUERY: To obtain all entities in GLO namespace

```
prefix itops: <http://www.ibm.com/ITOPS#>
prefix glo: <http://www.ibm.com/GLO#>

SELECT DISTINCT ?entity ?entityLabel WHERE {
            ?entity glo:subConceptOf glo:DomainEntity .
            FILTER(STRSTARTS(STR(?entity), "http://www.ibm.com/GLO#"))
            ?entity rdfs:label ?entityLabel .
 } ORDER BY ?entity
```

QUERY: to count all statements in the graph

```
SELECT DISTINCT (COUNT(*) as ?triples) WHERE
                        { ?s ?p ?o . }
```

QUERY: to count all statement minus those with metadata

```
prefix glo: <http://www.ibm.com/GLO#>
prefix itops: <http://www.ibm.com/ITOPS#>

SELECT DISTINCT (COUNT(*) as ?triples) WHERE {
    ?s ?p ?o .
   FILTER NOT EXISTS {
            ?p rdfs:subPropertyOf glo:topAnnotationProperty .
            ?p rdfs:subPropertyOf glo:topDataProperty .
             ?s rdfs:comment ?o .
            ?s rdfs:label ?o
     }
}
```

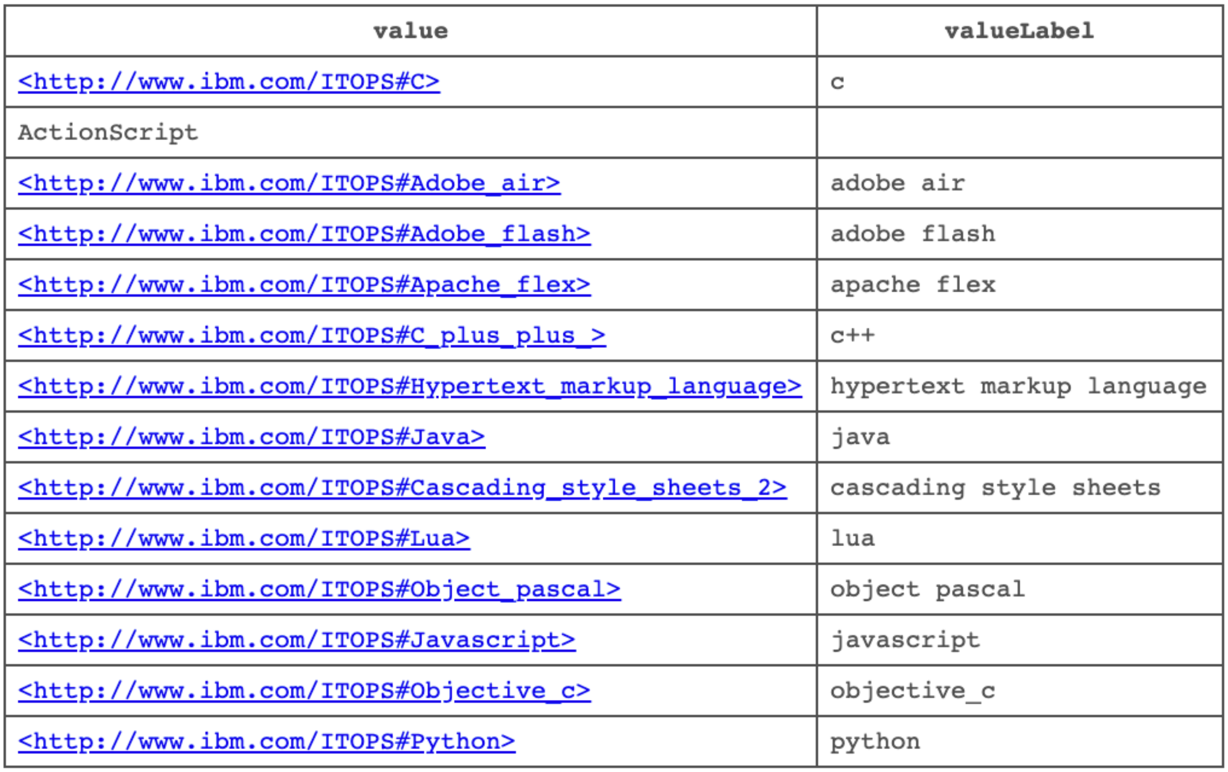
QUERY: to count all statement minus those with metadata

```
prefix itops: <http://www.ibm.com/ITOPS#>
prefix glo: <http://www.ibm.com/GLO#>

SELECT DISTINCT (COUNT(*) as ?triples) WHERE {
    ?s ?p ?o .
   FILTER NOT EXISTS {
            ?p rdfs:subPropertyOf glo:topAnnotationProperty .
            ?p rdfs:subPropertyOf glo:topDataProperty .
             ?s rdfs:comment ?o .
            ?s rdfs:label ?o
     }
```

QUERY: programming languages associated to “Adobe”

```
prefix glo: <http://www.ibm.com/GLO#>
prefix itops: <http://www.ibm.com/ITOPS#>

SELECT DISTINCT ?value ?valueLabel WHERE {
            ?item glo:subConceptOf itops:Software .
            { ?item rdfs:label ?itemLabel.
                        FILTER (CONTAINS(LCASE(?itemLabel), "adobe")) .
                        OPTIONAL { ?item glo:alias ?alias }
            }
            UNION
            {
                        ?item rdfs:label ?itemLabel .
                        ?item glo:alias ?alias .
                        FILTER (CONTAINS(LCASE(?alias), "adobe")) .
            }
            ?item itops:programming_language ?value .
            OPTIONAL { ?value rdfs:label ?valueLabel . }
 }
```

The result:



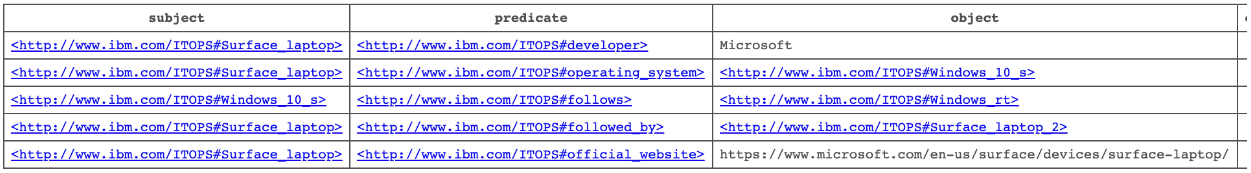
value	valueLabel
<http://www.ibm.com/ITOPS#C>	c
ActionScript	
<http://www.ibm.com/ITOPS#Adobe_air>	adobe air
<http://www.ibm.com/ITOPS#Adobe_flash>	adobe flash
<http://www.ibm.com/ITOPS#Apache_flex>	apache flex
<http://www.ibm.com/ITOPS#C_plus_plus_>	c++
<http://www.ibm.com/ITOPS#Hypertext_markup_language>	hypertext markup language
<http://www.ibm.com/ITOPS#Java>	java
<http://www.ibm.com/ITOPS#Cascading_style_sheets_2>	cascading style sheets
<http://www.ibm.com/ITOPS#Lua>	lua
<http://www.ibm.com/ITOPS#Object_pascal>	object pascal
<http://www.ibm.com/ITOPS#Javascript>	javascript
<http://www.ibm.com/ITOPS#Objective_c>	objective_c
<http://www.ibm.com/ITOPS#Python>	python

QUERY: Two hop graph (productive nodes)

Starting with itops:Laptop_model

```
prefix glo: <http://www.ibm.com/GLO#>
prefix ITOPS: <http://www.ibm.com/ITOPS#>

CONSTRUCT {
        ?x0 ?p0 ?x1 .
        ?x0 ?p1 ?x2 .
        ?x2 ?p2 ?x3 .
 }
WHERE {
        VALUES ?x0SUP {
        <http://www.ibm.com/ITOPS#Laptop_model>
         }
        VALUES ?x3SUP {
        <http://www.ibm.com/ITOPS#ITEntity>
        }

        ?x0 glo:subConceptOf ?x0SUP .
?       x3 glo:subConceptOf ?x3SUP .
        #Only domain properties
        ?p0 rdfs:subPropertyOf <http://www.ibm.com/GLO#domainEntityObjProp> .
        FILTER (?p0 != <http://www.ibm.com/GLO#domainEntityObjProp> ).
        ?p1 rdfs:subPropertyOf <http://www.ibm.com/GLO#domainEntityObjProp> .
        FILTER (?p1 != <http://www.ibm.com/GLO#domainEntityObjProp> ).
        ?p2 rdfs:subPropertyOf <http://www.ibm.com/GLO#domainEntityObjProp> .
        FILTER (?p2 != <http://www.ibm.com/GLO#domainEntityObjProp> ).
        ?x0 ?p0 ?x1 .
        ?x0 ?p1 ?x2 .
        ?x2 ?p2 ?x3 .
}
```

The results:



subject	predicate	object
<http://www.ibm.com/ITOPS#Surface_laptop>	<http://www.ibm.com/ITOPS#developer>	Microsoft
<http://www.ibm.com/ITOPS#Surface_laptop>	<http://www.ibm.com/ITOPS#operating_system>	<http://www.ibm.com/ITOPS#Windows_10_s>
<http://www.ibm.com/ITOPS#Windows_10_s>	<http://www.ibm.com/ITOPS#follows>	<http://www.ibm.com/ITOPS#Windows_rt>
<http://www.ibm.com/ITOPS#Surface_laptop>	<http://www.ibm.com/ITOPS#followed_by>	<http://www.ibm.com/ITOPS#Surface_laptop_2>
<http://www.ibm.com/ITOPS#Surface_laptop>	<http://www.ibm.com/ITOPS#official_website>	https://www.microsoft.com/en-us/surface/devices/surface-laptop/