# COMP474-6741 Project Assignment #1

Richard Nguyen Arman Jahanpour Wei Lien Huang Wasim Boughattas

April 2, 2022

## 1 Competency Questions

The competency questions focused on how the agent would be able to assess relations between the student class and the university class as well as testing various types of SPARQL queries. Including the first 3 that were mentioned in the handout, they are as follows:

1. What is course [SUBJECT][COURSE NUMBER], that is offered by [UNIVERSITY], about?
2. Which topics is [STUDENT FIRSTNAME] [STUDENT LASTNAME] competent in?
3. Which subjects at [UNIVERSITY] teach [TOPIC]?
4. What are all the courses with a [SUBJECT] subject at [UNIVERSITY]?
5. How many students are enrolled in each course that is offered by [UNIVERSITY]?
6. What courses are worth [CREDITS] credits at [UNIVERSITY]?
7. What are the topics of [SUBJECT][COURSE NUMBER] at [UNIVERSITY]?
8. Which students have retaken the same course at least [COUNT]?
9. Which students failed [SUBJECT][COURSE NUMBER] at [UNIVERSITY]?
10. What are the [REQUIREMENT STATUS] readings' [TITLE] and [URL] in [SUBJECT] [COURSE NUMBER] offered by [UNIVERSITY] in lecture [NUMBER]?

## 2 Vocabulary

The vocabulary and schema of the knowledge base were modeled by following standard W3C technologies: RDF and RDFS, by utilizing existing vocabularies [1]: Dublin Core, DBpedia, foaf, owl, vcard, wikidata, xsd, and vivo and were extended using developed classes and properties [1][2][3][4][5][6][7][8][9][10]. Table 1 highlights the classes and properties used to model the information [2]. Existing vocabularies where utilized in the modeling of the schema for properties and classes wherever possible as they're well maintained, updated regularly, commonly used, and link the data to existing knowledge graphs.

[1]For more details regarding classes and properties used from existing vocabularies refer to figure 1 in the appendix.

[2]For more details regarding developed classes and properties refer to the schema.

| Information | Class/Property |
|---|---|
| Universities | vivo:University |
| University name | rdfs:label |
| Link | owl:sameAs |
| Courses | vivo:Course |
| Course name | vivo:Title |
| Course subject | vivo:hasSubjectAre |
| Course number | vivo:identification |
| Course credits | vivo:CourseCredits |
| Course description | vivo:description |
| Course outline | focu:outline |
| Lectures | focu:lecture |
| Lecture number | bibo:number |
| Lecture name | vivo:Title |
| Lecture content | vivo:contains |
| Slides | focu:slide |
| Worksheets | focu:worksheet |
| Readings | focu:readings |
| Other material | focu:otherMaterial |
| Topics | focu:topic |
| Provenance information | focu:source |
| Students | focu:vivoStudent |
| Student name(first, last) | foaf:givenName, foaf:familyName |
| Student ID number | vivo:identification |
| Student email | foaf:mbox |
| Student completed courses | focu:completedCourse |
| Student competencies | focu:hasExpertise |

Table 1: Information and Vocabulary

## 3 Knowledge Base Construction

The data for courses was taken from Concordia open data website [11] and the data for universities was taken from DBpedia. As for other files (e.g., outline, slides, worksheets, ... of previous courses) local copies attained from course website and Moodle were used. To populate the knowledge base, for each information class: universities, courses, lectures, topics, and students a getRDF script was developed to generate related triples. The getRDF script for universities, courses, lectures, and students generates the triples automatically while the getRDF script for topics is hard coded to generate triples for topics covered in COMP 474 and COMP 6721 courses [3].

[3]The triples are hard coded for phase 1 as mentioned in the instructions extracting information about topics requires NLP techniques that will be part of phase 2

# 4 Knowledge Base Statistics

In addition, we provided queries and their output for statistics about our knowledge base, such as the total number of triples and the number of course URIs:

| triple_count | courseURI_count |
|---|---|
| 309818 | 7154 |

# 5 Queries

For translating the competency questions into queries, the values that are surrounded by square braces are given values for the queries. These would be used as starting points for the queries, eventually these will be part of the user's input which will change depending on the request without having to rewrite the entire query. As for the return values, these depended on what the question was. In general, the question was analyzed to see which class would contain the desired values then using the input values to form a series of triples that would link to them. Example outputs of the queries are as follows, note that for some of the queries only a small sample of their output is shown and due to the random generated nature of the knowledge base, the outputs shown are for a certain instance of the knowledge base:

| courseName | courseNumber | title | courseDesc |
|---|---|---|---|
| COMP | 474 | Intelligent Systems | Rule-based exp... |

Query 1: What is COMP 474, that is offered by Concordia University, about?

| expertise |
|---|
| isthecanonical URI |
| Nowthesamebookin French |
| rdf |
| ... |

Query 2: Which topics is Bryan Trenae competent in?

| title | subjectArea |
|---|---|
| Theoretical Problems in Religion and Culture | RELI |
| The New Imperialism | SOCI |
| Peoples and Cultures of Sub-Saharan Africa | ANTH |
| ... | ... |

Query 3: Which subjects at Concordia University teach anthropology?

| title | subject | courseNumber |
|---|---|---|
| SELECTED TOPICS IN SOFTWARE | COMP | 749 |
| Techniques in Symbolic Computation | COMP | 367 |
| Design and Analysis of Algorithms | COMP | 465 |
| ... | ... | ... |

Query 4: What are all the courses with a COMP subject at Concordia University?

| ?subject | ?catalog | ?count |
|---|---|---|
| ACCO | 220 | 5 |
| ACCO | 230 | 4 |
| ACCO | 240 | 6 |
| ACCO | 310 | 4 |
| ... | ... | ... |

Query 5: How many students are enrolled in each course that is offered by Concordia University?

| title | subjectCode | courseNum | credit |
|---|---|---|---|
| NONLINEAR SYSTEMS | ENGR | 6141 | 4.0 |
| TUTORIAL IN INF SYS/... | ACCO | 603 | 4.0 |
| CAPSTONE AEROSPACE ... | AERO | 490 | 4.0 |
| ... | ... | ... | ... |

Query 6: What courses are worth 4.0 credits at Concordia University?

| expertise |
|---|
| isthecanonical URI |
| Nowthesamebookin French |
| rdf |
| ... |

Query 7: What are the topics of COMP 474 at Concordia University?

| subjectArea | catalog | studentID | nbTimesTaken |
|---|---|---|---|
| AHSC | 330 | 43248687 | 2 |
| ANTH | 270 | 41103045 | 2 |
| ANTH | 345 | 47855211 | 2 |
| ... | ... | ... | ... |

Query 8: Which students have retaken the same course at least twice?

| studentID | courseName | courseNumber | grade |
|---|---|---|---|
| 43952942 | SOCI | 336 | F |

Query 9: Which students have failed SOCI 336 at Concordia University?

| requirement | title | website |
|---|---|---|
| ...data#required | [Yu14,Chapters1,2]... | https://concordia... |
| ...data#supplemental | [Wor14](RDFPrimer)... | http://www.w3.org... |
| ...data#supplemental | [RN10,Chapter12]... | https://concordia... |
| ...data#supplemental | Graphdatabases... | https://www.youtube. |

Query 10: What are the readings for COMP 474 offered by Concordia University in lecture 2?

# 6 References


[1] "RDF 1.1 Concepts and Abstract Syntax." [Online]. Available: https://www.w3.org/TR/rdf11-concepts/. [Accessed: Mar. 22, 2022]

[2] "RDF Schema 1.1." [Online]. Available: https://www.w3.org/TR/rdf-schema/. [Accessed: Mar. 22, 2022]

[3] "DCMI Schemas." [Online]. Available: https://www.dublincore.org/schemas/. [Accessed: Mar. 22, 2022]

[4] "Home," DBpedia Association. [Online]. Available: https://www.dbpedia.org/. [Accessed: Mar. 22, 2022]

[5] "FOAF Vocabulary Specification." [Online]. Available: http://xmlns.com/foaf/spec/. [Accessed: Mar. 22, 2022]

[6] "OWL Web Ontology Language Overview." [Online]. Available: https://www.w3.org/TR/owl-features/. [Accessed: Mar. 22, 2022]

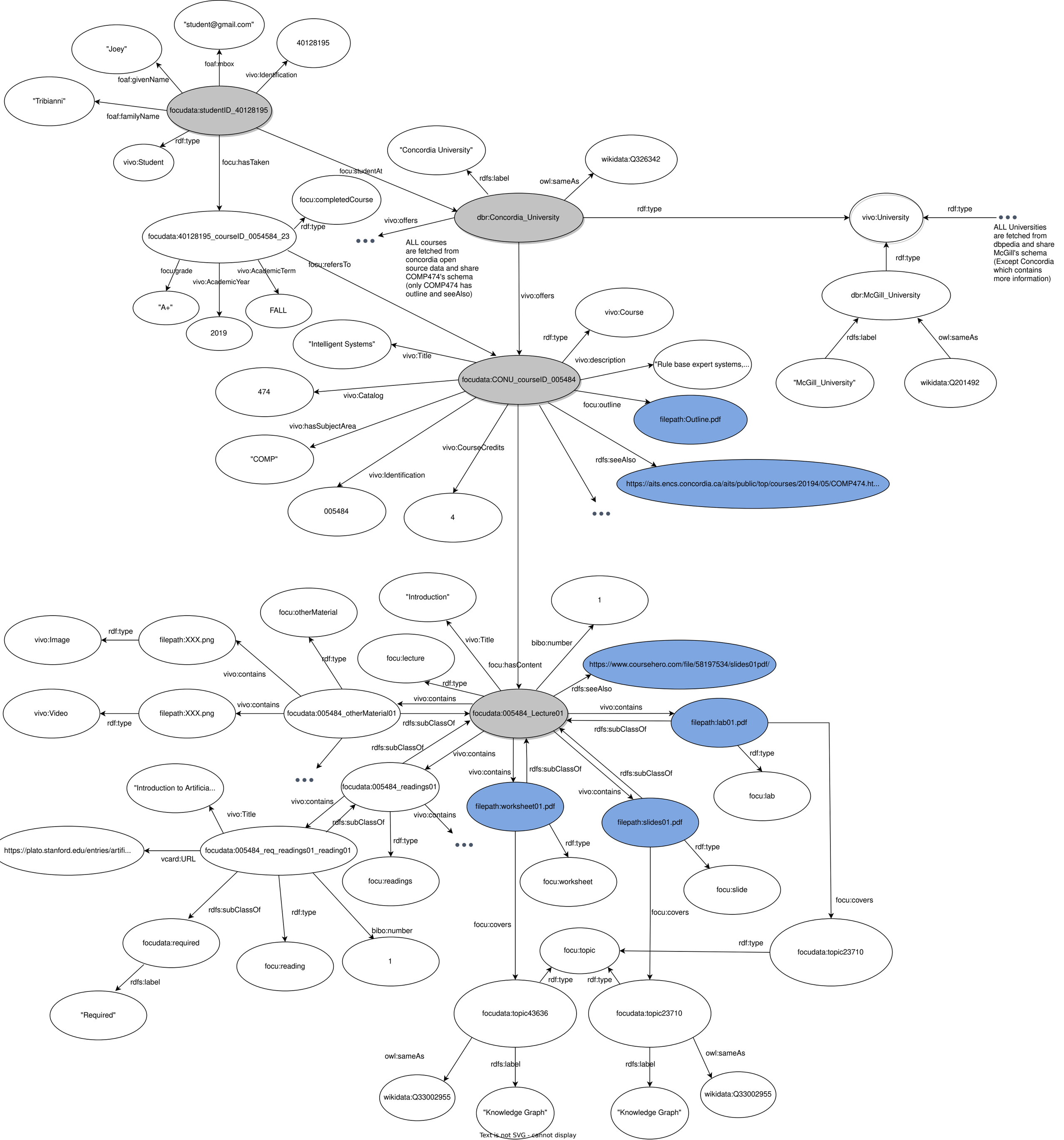
[7] "vCard Ontology - for describing People and Organizations." [Online]. Available: https://www.w3.org/TR/vcard-rdf/. [Accessed: Mar. 22, 2022]

[8] "Wikidata." [Online]. Available: https://www.wikidata.org/wiki/Wikidata:Main_Page. [Accessed: Mar. 22, 2022]

[9] "W3C XML Schema Definition Language (XSD) 1.1 Part 1: Structures." [Online]. Available: https://www.w3.org/TR/xmlschema11-1/. [Accessed: Mar. 22, 2022]

[10] "Linked Open Vocabularies (LOV)." [Online]. Available: https://lov.linkeddata.es/dataset/lov/vocabs/vivo. [Accessed: Mar. 22, 2022]

[11] "Opendata - Administrative module." [Online]. Available: https://opendata.concordia.ca/datasets/. [Accessed: Mar. 22, 2022]


"student@gmail.com"
"Joey"
40128195
foaf:mbox
foaf:givenName
vivo:Identification
"Tribianni"
focudata:studentID_40128195
foaf:familyName
rdf:type
vivo:Student
focu:hasTaken
focu:studentAt
"Concordia University"
rdfs:label
wikidata:Q326342
owl:sameAs
focu:completedCourse
rdf:type
vivo:offers
dbr:Concordia_University
rdf:type
vivo:University
rdf:type
ALL Universities are fetched from dbpedia and share McGill's schema (Except Concordia which contains more information)
focudata:40128195_courseID_0054584_23
ALL courses are fetched from concordia open source data and share COMP474's schema (only COMP474 has outline and seeAlso)
focu:refersTo
focu:grade
vivo:AcademicTerm
vivo:AcademicYear
"A+"
FALL
2019
vivo:offers
vivo:Course
rdf:type
vivo:description
"Rule base expert systems,...
"Intelligent Systems"
vivo:Title
focudata:CONU_courseID_005484
474
vivo:Catalog
focu:outline
filepath:Outline.pdf
vivo:hasSubjectArea
"COMP"
vivo:CourseCredits
rdfs:seeAlso
https://aits.encs.concordia.ca/aits/public/top/courses/20194/05/COMP474.ht...
vivo:Identification
005484
4
dbr:McGill_University
rdfs:label
owl:sameAs
"McGill_University"
wikidata:Q201492
"Introduction"
1
focu:otherMaterial
vivo:Image
rdf:type
filepath:XXX.png
vivo:Title
bibo:number
focu:lecture
https://www.coursehero.com/file/58197534/slides01pdf/
vivo:contains
rdf:type
focu:hasContent
rdfs:seeAlso
vivo:Video
filepath:XXX.png
vivo:contains
focudata:005484_otherMaterial01
focudata:005484_Lecture01
vivo:contains
filepath:lab01.pdf
rdfs:subClassOf
rdfs:subClassOf
rdf:type
rdfs:subClassOf
vivo:contains
vivo:contains
rdfs:subClassOf
rdfs:subClassOf
"Introduction to Artificia...
focudata:005484_readings01
filepath:worksheet01.pdf
filepath:slides01.pdf
focu:lab
vivo:contains
vivo:Title
rdfs:subClassOf
vivo:contains
rdf:type
rdf:type
rdf:type
https://plato.stanford.edu/entries/artifi...
vcard:URL
focudata:005484_req_readings01_reading01
focu:readings
focu:worksheet
focu:slide
focu:covers
focu:covers
rdfs:subClassOf
rdf:type
bibo:number
focudata:required
focu:reading
1
focu:topic
rdf:type
focudata:topic23710
rdfs:label
"Required"
rdf:type
rdf:type
focudata:topic43636
focudata:topic23710
owl:sameAs
rdfs:label
rdfs:label
owl:sameAs
wikidata:Q33002955
"Knowledge Graph"
"Knowledge Graph"
wikidata:Q33002955
Text is not SVG - cannot display