# COMP474-6741 Project Assignment #1

Richard Nguyen          Arman Jahanpour          Wei Lien Huang          Wasim Boughattas

March 22, 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 [COURSE NAME][COURSE NUMBER] about?

2. Which topics is [STUDENT FIRSTNAME] [STUDENT LASTNAME] competent in?

3. Which courses at [UNIVERSITY] teaches [TOPIC].

4. What are all the courses for [COURSE NAME].

5. How many students are registered for [COURSE ID]

6. What courses are worth [CREDITS] credits?

7. What are the topics of [COURSE NAME][COURSE NUMBER]

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

9. Which students failed [COURSE NAME][COURSE NUMBER] and did not retake it?

10. What are the [REQUIREMENT STATUS] readings' [TITLE] and [URL] in [COURSE SUBJECT AREA] [COURSE CATALOG] 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.

| 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].

## 4 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

---

[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.

[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

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 expert systems... |

Table 2: What is COMP 474 about?

| label | expertises |
|---|---|
| State Space Search | ...courseID_040353_topic_1 |
| Alpha-Beta Pruning | ...courseID_040353_topic_2 |
| Artificial Neural Networks | ...courseID_040353_topic_7 |
| ... | ... |

Table 3: Which topics is Karra Walton competent in?

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

Table 4: Which courses at Concordia teaches anthropology?

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

Table 5: What are all the courses for COMP

| ?id | ?title | ?count |
|---|---|---|
| 47892 | Promote your Research using Open Access | 4 |

Table 6: How many students are registered for courseID 47892?

| title | subjectCode | courseNumber | credit |
|---|---|---|---|
| NONLINEAR SYSTEMS | ENGR | 6141 | 4 |
| TUTORIAL IN INF SYS/INT.CONT | ACCO | 603 | 4 |
| CAPSTONE AEROSPACE ENGINEERING DESIGN PROJECT | AERO | 490 | 4 |
| ... | ... | ... | ... |

Table 7: What courses are worth 4.0 credits?

| label |
|---|
| Artificial Neural Network |
| Intelligent Agent |
| SPARQL |
| ... |

Table 8: What are the topics of COMP 474?

| subjectArea | catalog | studentID | nbTimesTaken |
|---|---|---|---|
| CEPS | 1012EO | 49182836 | 2 |
| CECR | 1102 | 47266970 | 2 |
| CEBD | 1300 | 49237110 | 2 |
| ... | ... | ... | ... |

Table 9: Which students have retaken the same course at least twice?

| studentID | courseName | courseNumber | grade |
|---|---|---|---|
| 44067924 | CERA | 452 | F |

Table 10: Which students has failed CERA 452 and did not retake it?

| uniName | subjectArea | catalog | lecNum |
|---|---|---|---|
| Concordia University | COMP | 474 | 2 |
| Concordia University | COMP | 474 | 2 |
| Concordia University | COMP | 474 | 2 |
| Concordia University | COMP | 474 | 2 |

Table 11: What are the required readings for COMP 474 offered by Concordia in lecture 2?

| 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. |

Table 12: What are the readings for COMP 474 offered by Concordia in lecture 2? (Cont.)

# 5 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]

Nodes and relationships:

- "student@gmail.com"
- "Joey"
- 40128195
- "Tribianni"
- focudata:studentID_40128195
  - foaf:mbox → "student@gmail.com"
  - foaf:givenName → "Joey"
  - vivo:identification → 40128195
  - foaf:familyName → "Tribianni"
  - rdf:type → vivo:Student
  - focu:hasTaken → focudata:40128195_courseID_0054584
  - focu:studentAt → dbr:Concordia_University
- vivo:Student
- vivo:AcademicTerm
- focudata:40128195_0054584_term
  - rdf:type → vivo:AcademicTerm
  - vivo:contains → "SUMMER-1-2022"
  - vivo:contains → "FALL-2019"
- "SUMMER-1-2022"
- "FALL-2019"
- focudata:40128195_courseID_0054584
  - focu:history → focudata:40128195_0054584_term
  - rdf:type → focu:completedCourse
  - focu:refersTo → focudata:courseID_005484
  - focu:hasHighestGrade → "A+"
- focu:completedCourse
- "A+"
- "Concordia_University"
- wikidata:Q326342
- dbr:Concordia_University
  - rdfs:label → "Concordia_University"
  - owl:sameAs → wikidata:Q326342
  - rdf:type → vivo:University
  - vivo:offers → focudata:courseID_005484
- vivo:University
  - rdf:type → ...
  - ALL Universities are fetched from dbpedia and share McGill's schema (Except Concordia which contains more information)
- ALL courses are fetched from concordia open source data and share COMP474's schema (only COMP474 has outline and seeAlso)
- dbr:McGill_University
  - rdf:type → vivo:University
  - rdfs:label → "McGill_University"
  - owl:sameAs → wikidata:Q201492
- "McGill_University"
- wikidata:Q201492
- "Knowledge Graph"
- wikidata:Q33002955
- focu:topic
- focudata:courseID_5484_topic_1
  - rdfs:label → "Knowledge Graph"
  - owl:sameAs → wikidata:Q33002955
  - rdf:type → focu:topic
  - focu:hasExpertise → focudata:40128195_courseID_0054584
  - focu:coveredIn → focudata:courseID_005484
  - focu:source → https://aits.encs.concordia.ca/aits/public/top/courses/20194/05/COMP474.ht...
- vivo:Course
- "Intelligent Systems"
- 474
- "COMP"
- 005484
- 4
- "Rule base expert systems,..."
- focudata:courseID_005484
  - rdf:type → vivo:Course
  - vivo:Title → "Intelligent Systems"
  - vivo:description → "Rule base expert systems,..."
  - vivo:Catalog → 474
  - vivo:hasSubjectArea → "COMP"
  - vivo:CourseCredits → 4
  - vivo:Identification → 005484
  - focu:outline → filepath:Outline.pdf
  - rdfs:seeAlso → https://aits.encs.concordia.ca/aits/public/top/courses/20194/05/COMP474.ht...
  - focu:hasContent → focudata:005484_Lecture01
- filepath:Outline.pdf
- https://aits.encs.concordia.ca/aits/public/top/courses/20194/05/COMP474.ht...
- "Introduction"
- 1
- https://www.coursehero.com/file/58197534/slides01pdf/
- focu:otherMaterial
- focu:lecture
- vivo:Image
- filepath:XXX.png (rdf:type → vivo:Image)
- vivo:Video
- filepath:XXX.png (rdf:type → vivo:Video)
- focudata:005484_otherMaterial01
  - rdf:type → focu:otherMaterial
  - vivo:contains → filepath:XXX.png
  - rdfs:subClassOf → focudata:005484_Lecture01
- focudata:005484_Lecture01
  - vivo:Title → "Introduction"
  - bibo:number → 1
  - rdf:type → focu:lecture
  - rdfs:seeAlso → https://www.coursehero.com/file/58197534/slides01pdf/
  - vivo:contains → focudata:005484_readings01
  - vivo:contains → filepath:worksheet01.pdf
  - vivo:contains → filepath:slides01.pdf
- "Introduction to Artificia..."
- https://plato.stanford.edu/entries/artifi...
- focudata:005484_req_readings01_reading01
  - vivo:Title → "Introduction to Artificia..."
  - vcard:URL → https://plato.stanford.edu/entries/artifi...
  - rdfs:subClassOf → focudata:required
  - rdf:type → focu:reading
  - bibo:number → 1
- focudata:005484_readings01
  - rdfs:subClassOf → focudata:005484_Lecture01
  - vivo:contains → focudata:005484_req_readings01_reading01
  - rdf:type → focu:readings
- focudata:required
- focu:reading
- focu:readings
- filepath:worksheet01.pdf
  - rdfs:subClassOf → focudata:005484_Lecture01
  - rdf:type → focu:worksheet
- filepath:slides01.pdf
  - rdfs:subClassOf → focudata:005484_Lecture01
  - rdf:type → focu:slide
- focu:worksheet
- focu:slide