COMP474-6741 Project Assignment #1

Richard Nguyen Arman Jahanpour Wei Lien Huang Wasim Boughattas

April 12, 2022

1 Competency Questions Part I

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 FIRST NAME] [STUDENT LAST NAME] 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
Outline	focu:outline
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 and Population

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. To generate topics triples, course material (slides, worksheets, labs, and outline) were converted to plain text using pdfplumber library [12]. Next, DBpedia Spotlight is run over the converted documents for entity linking from which a list of entities and their DBpedia URI is extracted [13]. Lastly the aforementioned list is passed through a spaCy filter to extract named entities [14]. The filter is designed to select entities with specific POS tags obtained from spaCy glossary [15].

4 Knowledge Base Statistics

In addition, we provided queries and their output for statistics about our knowledge base, including the total number of triples, the total number of courseURIs, the number of distinct topics, and the number of topic/course instances:

triples	courseURIs	distinct_topics	topic_instances
309241	7154	329	817

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

5.1 Part I

NB— Implemented in Chatbot (✓)

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

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

expertise
Heidelberg
Stuart Russell
Diffbot
...

Part I: Query 2: Which topics is Trenae Bryan competent in?

courseName	subjectArea	courseNum	frequency
Intelligent Systems	RELI	474	48
INTRO TO A.I.	SOCI	6721	4

Part I: Query 3: Which subjects at Concordia University teach DBpedia?

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

Part I: 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
...	...	...

Part I: 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
...	...	...	...

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

expertise
Heidelberg
Stuart Russell
Diffbot
...

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

subjectArea	catalog	firstName	lastName	studentID	nbTimesTaken
AHSC	330	Breia	Crowell	43248687	2
ANTH	270	Sanya	Ray	41103045	2
ANTH	345	Sinead	Thomas	47855211	2
...	...	...	...	...	...

Part I: Query 8: Which students have retaken the same course at least 2 times?

firstName	lastName	studentID	courseName	courseNumber	grade
Mariela	Morrow	43952942	SOCI	336	F

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

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

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

5.2 Part II

5.2.1 Knowledge Base Population

NB— Not Implemented in Chatbot (✗)

subject	catalog	topicLabel	dbpediaURI	eventURI	resourceURI
COMP	474	Heidelberg	...Heidelberg	...	...
COMP	474	... Russell	..._Russell	...	...
...	...	...	...	...	...
COMP	6721	Centroid	.../Centroid	...	...
...	...	...	...	...	...

Part II: Query 1: For a course c, list all covered topics t, printing out their English labels and their DBpedia URI, together with the course event URI (e.g., 'lab3') and resource URI (e.g., 'slides10') where they appeared.

dbpediaURI	topicLabel	catalog	subject	topicCount
.../Concordia_University	Concordia	474	COMP	75
.../DBpedia	DBpedia	474	COMP	36
.../English_language	English language	474	COMP	24
...	...	...	...	...
.../Artificial_intelligence	AI	6721	COMP	18
.../Natural_language_processing	Natural Language ...	474	COMP	15
...	...	...	...	...

Part II: Query 2: For a given topic t (DBpedia URI), list all courses where they appear, together with a count, sorted by frequency

topicLabel	courseURI	eventURI	resourceURI
Data Publishing	...#courseID_005484	.../data#005484_Lecture05	.../slides05.pdf
Weak AI	...#courseID_040353	.../data#040353_Lecture01	.../slides01.pdf
...	...	...	...

Part II: Query 3: For a given topic t, list the precise course URI, course event URI and corresponding resource URI where the topic is covered (e.g., "NLP" is covered in COMP474 → Lecture 10 → Lab 10 → Lab Notes)

5.2.2 University Chatbot

NB— Implemented in Chatbot (✓)

Some of the queries coincide with existing queries in part I:

1. see queries 1-10 in part I
2. "What is the [course] about?" : see query 1 in part I
3. "Which topics are covered in [course event]?" : See below

topicLabel	resourceURI
XOR	/worksheet05.pdf
neuron	/worksheet05.pdf
Perceptron	/worksheet05.pdf
Delta	/worksheet05.pdf
Alison	/worksheet05.pdf

Part II: Query 4: Which topics are covered in worksheet 5 of COMP 6721?

4. "Which courses cover [Topic]?" : see query 3 in part I

6 Rasa Chatbot

For designing the chatbot, we used the competency questions that were associated with each query. That way, we can see where the chatbot should extract the variables. This is represented by the words surrounded by square brackets as seen in Section 1.The variables would be extracted using Rasa's entity system to pull the text from input and classify them. Once the variables are extracted and stored in the chatbot's memory, which is done using the slots, it would place those into the appropriate query. We provided structure of the queries to the bot in its custom action classes and it just needs to substitute the variables. Afterwards, the bot makes a request to the Fuseki server, that contains our Knowledge Base, with the completed query and receives a response from the server. To make the output more human-readable, as the response is in JSON format, we extract the values from it and lay it out as a text list which the chatbot will display onto the command terminal. As for dealing with exception cases, the bot has a few responses built in to handle them by replying to a message to communicate the issue with the user.

For the sample responses from the chatbot for the questions from the user, please see the appendix at the end. Note that for some of the responses, only a portion of it is shown.

7 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]

[12] "pdfplumber." https://github.com/jsvine/pdfplumber/ [accessed Apr. 11, 2022].

[13] "DBpedia Spotlight - Shedding light on the web of documents." https://www.dbpedia-spotlight.org/ [accessed Apr. 11, 2022].

[14] "spaCy · Industrial-strength Natural Language Processing in Python." https://spacy.io/ [accessed Apr. 11, 2022].

[15] M. Honnibal, I. Montani, S. Van Landeghem, and A. Boyd, spaCy: Industrial-strength Natural Language Processing in Python. 2020. Accessed: Apr. 11, 2022. [Online]. Available: https://github.com/explosion/spaCy/blob/d4196a62f198f0ec32239b238f32421bbb6eb942/spacy/glossary.py

[15] "DBpedia Spotlight - Shedding light on the web of documents." https://www.dbpedia-spotlight.org/ [accessed Apr. 11, 2022].

8 Appendix

See next page

9 Graph

See final page