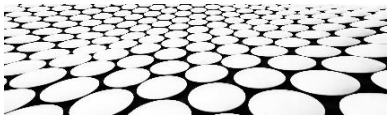
CSI5101 Knowledge Representation



Assignment 2

Ontologies and Logics

GOALS

The purpose of this assignment is to further familiarize yourself with the different concepts seen during our exploration of ontologies and logics. You will further look at ontology design, linked-open data, Google's knowledge graphs, propositional and predicate logic, as well as fuzzy logic and description logics.

SUBMISSION DEADLINE

- Assignment released: Tuesday, February 6
- Assignment submission ends: Monday, February 26, midnight (11:59pm)*

**Because of the reading week, the deadline is after 3 weeks instead of 2.*

SUBMISSION METHOD / CONTENT

- You can work alone or in groups of 2 (your choice). If you work in groups, both students should submit the same report. The assignment is the same whether you work by yourself or in a team.
- A link in Brightspace is provided for Assignment 2.
- Submit a PDF report containing your answers.
- Your report should contain a title page, answer to each question clearly identified. It should be easy to go through.

INSTRUCTIONS

Many questions ask you to explain and provide examples to illustrate your explanation. It is appropriate to find information online, but you MUST cite your sources. For any example, illustration or explanation that does not come from your own thinking, you should say what source you used. It is important to acknowledge where you found your information. And it shows that you searched. For your answers, try to find good/reliable sources.

Most questions (except Q5) are open questions. **For open questions, there are no right or wrong answers in terms of content.** BUT, there are "well thought-out, well argumented and well written answers" versus "quickly written, poorly argumented and not clear answers". Many questions related to Ontologies require that you explore resources, so your answers should show that you actually spent time to explore.

Try to provide good information, good examples, but still be concise. Provide the sources after each question, DO NOT state them all at the end.

EVALUATION

- This assignment is worth 10%.
- There are 7 questions, and 60 points total.
- Although very open in nature, each question requires thinking, reading, investigation of resources and analysis. It will be expected that each student/group does take the time to provide answers that are well thought-out and justified.
- Furthermore, attention will be given to:
    - Clarity. Student is able to express his/her ideas clearly. Good argumentation. Good examples chosen.
    - Conciseness. Student is able to express his/her ideas in a concise way.
    - Information gathering/analysis: Student found good/interesting information in articles/blogs/videos and is able to analyze this information.
    - Effort: Student put effort into answering the question.
    - Sources provided: Student cites his/her sources (article, blog, video, image), when used.

QUESTIONS

- Something is not clear? You can ask your questions in the Brightspace Forum within the Assignment topic.

Ontologies

Q1 – Ontology Design (12 points)

In Ontology Design, an important step is to provide "Competency Questions", so figuring out what is in-scope and out-of-scope for the ontology that we design. One other principle is to not reinvent the wheel and look for existing ontologies that could correspond to our need.

Assume you need to design a small ontology about the University domain which would show classes and properties, as well as a small KB (Knowledge Bases) which would instantiate some classes defined in the ontology.

Assume the following in-scope competency questions:

- What is the list of departments at the University of Ottawa?
- What are the courses offered in the department of chemistry?
- What are the scholarships available for international students?

Assume the following out-of-scope question:

- What is the average age of professors in the EECS department?

TO DO:

- Add 2 more in-scope competency questions. Add 2 more out-of-scope competency questions.
- Explore Schema.org *CollegeOrUniversity* class (they call them types) https://schema.org/CollegeOrUniversity:
    - Does it contain sufficient properties for your needs?
    - Notice that multiple inheritance is permitted, as *CollegeOrUniversity* is a subtype of *EducationalOrganization* which is a subtype of both *CivicStructure* and *Organization*. Why do you think that is appropriate (or not)?
    - What are some properties inherited from *EducationalOrganization* that seem strange to you to describe a University?
- What about this other university ontology (link) ?
    - Does it contain what you need? Explore this ontology a bit.
- Building on classes from the 2 above-mentioned ontologies, define your own small university ontology and KB containing the classes and relations to answer your 5 competency questions.
- You can write out your ontology by providing for each class: name, superclass, properties, and by providing for each property: domain, range.
- You can populate your small KB with only a few examples that could answer the competency questions.

Write your answer to contain all the information required and organized the way you want to best present your design and your exploration. Your answer can take a few pages if needed.

Q2 – Linked Open Data (LOD) in Biomedical domain (8 points)

Some domains, like the biomedical domain, have developed and adopted various ontologies. Furthermore, the ontologies are linked to each other, as promoted by the Linked Open Data philosophy.

If you start here: https://guides.lib.umich.edu/ontology/ontologies, you'll get a list of different ontologies. We'll focus on the Disease Ontology (https://disease-ontology.org/) .

Pick one disease (does not matter each one). You'll see links to ICD9, ICD10, MESH, SNOMEDCT, UMLS. These are all vocabularies or ontologies in the medical domain, and there are links from a particular disease to those other vocabularies.

TO DO:

- Define your starting point (what did you choose as disease).
- Explore the other ontologies linked to (ICD9, ICD10, MESH, SNOMEDCT, UMLS). What are those? Are they open source? Are they published by authorities? Are they ontologies, or are they lists or vocabularies (without a hierarchical backbone and properties)?
- From the disease you chose, what does each ontology provide as additional information that the Disease Ontology does not have? Give some examples for each one.

Write your answer to contain all the information required and organized the way you want to best present your exploration. The purpose is to discover the various medical ontologies and reflect on why there are so many, and also on how they are linked to each other. Take the time to explore. If possible, show information about your starting point (disease) from each ontology (ICD9, ICD10, MESH, SNOMEDCT, UMLS). Some of them require a login, so it won't be possible.

Q3 – Google Knowledge Graphs (GKG) and Search Engines (15 points)

As we discussed in class, although we do not yet have a Semantic Web as imagined by Tim Berners-Lee, a lot of the normalization (ontology development) is now used as annotation within web pages to help search engines locate specific structured information for which to provide a particular display. A lot of the normalized vocabulary is actually provided by Schema.org (https://schema.org/)

Here are two short blogs/articles about Schema.org and Google's Knowledge Graphs (GKG) and their use in Search Engines (for SEO), both written by Michal Pecánek, 2020

- Article about markup in web pages. What is Schema Markup? How to Use It for SEO?,
- Article about use of KG in Search Engines: Google's Knowledge Graph Explained: How It Influences SEO,

As a developer, you could also be interested by the Knowledge Graph API that Google provides, even if that is outside the scope of the current question.

Read the 2 articles suggested above, to be able to answer the following questions. The questions will also require that you navigate within Schema.org and explore classes and properties.

TO DO:

- Define what Knowledge Panels/Cards are and why they are used?
- Describe the relation between GKGs and Schema.org?
- In the second article, section 2 (*Use schema markup on your site*), look at the "sameAs" in the example. This "sameAs" link is fundamental in LOD. Why? What does it do?
- In that same section 2, it mentions to use the *Organization* markup. Take a look at the *Organization* class in Schema.org:
    - The superclass of *Organization* is *Thing*, and some properties are inherited from *Thing*. What are some of these properties? Do they seem relevant to *Organization*?
    - *Organization* can be the range for which properties?
    - Why is a type like *CreativeWork* used so much as a possible type for various properties?
    - Why is *Text* like "Text or PostalAddress" often used as possibly type for various properties? What are the advantages and disadvantages of using *Text* as a property type?
- Some classes don't have the same status in Schema.org, for example, look at VirtualLocation what's its status? What does that mean in the context of ontology development? Is Schema.org curated? If yes, by whom? Are they to be trusted?
- A *PostalAddress* is a subclass of *ContactPoint*, why would an ontology want to make that distinction?
- Go in *ListenAction*. Look at the examples (in JSON-LD, it's easy to read). They differentiate John listened to Pink. John listened to Star 101.3. How is such differentiation done?
- Let say you wanted to express that *Bora went to the National Art Center in Ottawa to hear a concert by the NAC Orchestra playing music by Beethoven*, what classes could be assigned to "Bora", "National Art Center", "NAC Orchestra", "Ottawa", "Beethoven". What are the properties between them? Inspired by the JSON-LD format shown in the *ListenAction* page

examples, represent the sentence above using that format. You can make-up classes if you do not find the appropriate ones in Schema.org.

Write your answer to contain all the information required and organized the way you want to best present your exploration. The overall purpose of this question is to get a better understanding of Schema.org, of its strength and limitations. Take the time to navigate within Schema.org to better appreciate this resource. You can be appreciative of all the effort put in the development of such a resource, but you can also be critical and present your views.

Logics

Q4 – Propositional logic and proofs (6 points)

This book, A Concise Introduction to Logic, by Craig DeLancey is freely available online.

https://milnepublishing.geneseo.edu/concise-introduction-to-logic/

Chapter 4 talks about proofs and is an easy read.

https://milnepublishing.geneseo.edu/concise-introduction-to-logic/chapter/4-proofs/

In 4.5, there are some exercises suggested.

e. Premises: **(S→¬Q)**, **(P→S)**, **¬¬P**. Show: **¬Q**.
f. Premises: **(T→P)**, **(Q→S)**, **(S→T)**, **¬P**. Show: **¬Q**.
g. Premises: **R**, **P**, **(P → (R → Q))**. Show: **Q**.
h. Premises: **((R→S)→Q)**, **¬Q**, **(¬(R→S)→V)**. Show: **V**.
i. Premises: **(P→(Q→R))**, **¬(Q→R)**. Show: **¬P**.
j. Premises: **(¬(Q→R)→P)**, **¬P**, **Q**. Show: **R**.
k. Premises: **P**, **(P→R)**, **(P→(R→Q))**. Show: **Q**.
l. Premises: **¬R**, **(S→R)**, **P**, **(P→(T→S))**. Show: **¬T**.
m. Premises: **P**, **(P→Q)**, **(P→R)**, **(Q→(R→S))**. Show: **S**.
n. Premises: **(P→(Q→R))**, **P**, **((Q→R)→¬S))**, **((T→V)→S)**. Show: **¬(T→V)**.

TO DO:

- Choose 4 of these (between e and n) that you prove using rules of inferences.
- Present each proof as we saw in class (e.g. slide 28):
    - State the premises (assigning a number to each one).
    - For each intermediate step (until you arrive at the conclusion), state what is derived (either inferred or equivalent) and how (which equivalence rule or which inference rule is used).

Q5 – Predicate logic (6 points)

Predicate logic has a certain expressive power. There are facts it can express and others it cannot. Its representation relies mostly on predicates (unary and binary) and on quantifiers (universal and existential).

TO DO:

Express the following statements in **predicate logic**. If you think the statement cannot be expressed in first-order logic, explain why.

- All graduate students have an undergraduate degree.
- Anyone with an undergraduate degree studied in at least one university.
- A university, at any particular year, deliver courses to many students.
- For a professor to teach a course, there must be a student attending it.
- Classmates are students taking the same course.
- A graduate student takes less courses than an undergraduate student.

Q6 – Description Logics (8 points)

You are asked to develop a description logic (DL) system for defining a recycling world.

T-Box

- Boxes are either Blue boxes or Black Boxes
- Blue box is a box that contains only glass or metal
- Black box is a box that contains only paper or cardboard
- If something is made of glass, it is a subclass of glass
- If something is made of cardboard, it is a subclass of cardboard
- Everything recyclable is either glass, metal, paper, cardboard.

A-Box (instances)

- A wine bottle (w) made of glass
- A newspaper (n) made of paper

TO DO:

- Represent in DL the ontology described above.
- Add at least 2 more rules in the T-Box.
- Add 5 more instances in the A-Box that would make the world "satisfiable".
- Give 2 examples of instances that you could add to make the world "unsatisfiable" and explain why. No need for formal proofs, just an explanation in words.

ATTENTION: I know we have not spent a lot of time on DL, but with the slide (X), you have enough to understand that classes are defined and how they are defined (through the use of restrictions on properties).

Q7 – Fuzzy Logic (5 points)

There are some industrial applications of fuzzy logic. A train in Japan was operated with fuzzy logic for a while (making the formalism more popular). Today's rice cookers and washing machine vendors claim they use a fuzzy logic technology…

TO DO:

- Find some information online about the current (or past) use of fuzzy logic for controlling systems (from train operations to farming). Cite some sources and summarize them.
- Do rice cookers and washing machine really use fuzzy logic? How?
- What about this Japanese train, what's its story?