UNIVERSITY OF ABERDEEN

University of Aberdeen

School of Natural and Computing Sciences

Department of Computing Science

MSc in Artificial Intelligence

2020 – 2021

Please read all the information below carefully
Assessment Item 2 of 2 Research Project – Individually Assessed (no teamwork)

Course: CS551J – Knowledge Representation and Reasoning **Research Project/Scientific Report**	***Note: This assessment accounts for 50% of your total mark of the course.***

Learning Outcomes

On successful completion of this component a student will have demonstrated competence in the following areas:

- Demonstration of understanding advanced principles and concepts of KRR.
- Creation of knowledge graph.
- Analysis and discussion of domain knowledge graph.
- Effective communication of complex concepts.

Information for Plagiarism and Conduct: Your submitted report may be submitted for plagiarism check (e.g., Turnitin). Please refer to the slides available at MyAberdeen for more information about avoiding plagiarism before you start working on the assessment. Please also read the following information provided by the university: https://www.abdn.ac.uk/sls/online-resources/avoiding-plagiarism/

In addition, please familiarise yourselves with the following document "code of practice on student discipline (Academic)": https://tinyurl.com/y92xgkq6

Assessment Tasks & Report Guidance

Your report must conform to the below structure and include the required content as outlined in each section. Each subtask has its own marks allocated. You must supply a written report, along with the corresponding code, containing all distinct sections/subtasks that provide a full critical and reflective account of the processes undertaken.

Overview

Knowledge Graph (KG) is a powerful knowledge representation and reasoning tool for linking structured and unstructured data together. For example, semantic knowledge graph combines structured and unstructured information on the Web so that computer systems can understand text.

Please read all the information below carefully

Consequently, knowledge graph enables search techniques and question answering. The aims of this assessment are to understanding knowledge representation and reasoning using KG, creating and encoding a text knowledge graph, and knowledge graph analysis. In this assessment, you will particularly focus on the Coronavirus (COVID-19) vaccines; for COVID-19 vaccine information, please refer to the NHS site "https://www.nhs.uk/conditions/coronavirus-covid-19/coronavirus-vaccination/coronavirus-vaccine/". Following these tasks, you will create a COVID-19 vaccine knowledge graph (COVID-VKG) from scratch and perform a basic analysis. Note that in your report, for each of the following tasks you must demonstrate the process how it was performed by providing graphs and code snippets and provide analysis if necessary.

Tasks:

1. Collect resources from Web articles/texts related to COVID-19 vaccines. You can use existing software such as Python tools to crawl Web pages related to this topic. (**10%**)
2. Extract texts from the crawled web pages (**5%**) and extract sentences relevant to COVID-19 vaccines from these texts; you may need to remove the noise sentences. (**5%**)
3. Given the extracted sentences, write a programme to extract the triples required for creating COVID-VKG. These triples can be represented as structured data which include not only the triples, but also the resources and the sentences. (**20%**)
4. As the data have been represented as structured data, use the principles of Resource Description Framework (RDF) to create the COVID-19 vaccine knowledge graph. Note that the entities extracted from text should be mapped to a vacabulary in a known ontology such as Dbpedia. If an extracted entity can be mapped to an identical entity in any known ontology, the uniform resource identifier (URI) should be used as a representative for the extracted entity; otherwise, a new URL is given to the entity. Represent the COVID-VKG by using N-Triples and Turtle serialisation methods. You should explicitly show some examples of your COVID-VKG in your report and describe how you have created the KG. (**40%**)
5. After the COVID-VKG has been created, use querying tools to extract knowledge from the graph. For example, this can be used for question answering to answer the following questions: What are the approved vaccines and thus are being used given a country? What are the side effects for a vaccine? You should explicitly explain how your programme answers these questions and the outcome of your programmes. (2**0%**)

Marking Criteria

- Quality of the report, including structure, clarity, and brevity (Academic writing skills)
- Communication skills (clear technical content)
- Critical approach (sound reasoning and evaluation)
- Depth and breadth of the evaluation (is my evaluation robust)

Please read all the information below carefully

Submission Instructions

You should submit a PDF version of your report/essay via MyAberdeen by 23:59 on **Friday 5th March**. The name of the PDF file should have the form "CS551J_Assessment2_<Your Student ID>". For instance, "CS551J_Assessment2 _4568985.pdf", where 4568985 is your student ID.

Please try to make your submission file less than 20MB as you may have issues when uploading large files to MyAberdeen.

Any questions pertaining to any aspects of this assessment, please address them to the course lecturer Dr. Mingjun Zhong, mingjun.zhong@abdn.ac.uk.