Semantic Web (CSE632)
Winter 2020
Assignment 5

Instructions

1. This assignment is not a group activity. Each student has to work on it by himself/herself and submit the assignment.
2. Plagiarism check and corresponding policies will be strictly enforced. Students can be selected at random for a "viva" on the assignment and if the responses are not satisfactory, they will get a 0 in the assignment.
3. SHACL constraints and reports can go into a folder named "shacl". Within that folder, all constraints should be in a "constraints" folder. All the reports should go into the "reports" folder. All the textual description of the reports can go into "txt-reports" folder. Each constraint should be in a separate file named Q1-#.ttl. Follow the same convention for the reports and txt-reports as well. There should only be one pdf file for Q2. Include the images, if any, into this pdf itself. Create a zip file with the name RollNo_HW5.zip and submit it.
4. If the instructions from points 3 are not followed, there will be a penalty of 10% reduction in the points.

Questions **Max points: 50**

1. Validate your Netflix movie RDF graph (put all the triples into one file) using the following SHACL constraints and generate a report in each case. Use any of the SHACL libraries listed below and submit the SHACL constraints and the reports. **15 pt**
    a. Place a constraint that every movie should have at least one director and it should be released in at least one country.
    b. Construct a shapes graph to check the constraint that a movie cannot belong to both comedy and romantic genre.
    c. Using a regular expression, place a constraint on the title of the movie that it can only have alphanumeric characters and space in the title.
    d. For any object of a triple that is an IRI, validate the type of the object and also make sure that all such objects do indeed have an IRI.
    e. Write any constraint (apart from the above four) to demonstrate closed and ignoredProperties

    Explain in 2-3 sentences the report of each SHACL constraint in plain English. **5 pt**

    SHACL Libraries
    - https://jena.apache.org/documentation/shacl/index.html
    - https://github.com/TopQuadrant/shacl/

- https://github.com/AKSW/RDFUnit/
- https://github.com/RDFLib/pySHACL
- https://github.com/weso/shaclex

2. Triple stores and property graph stores
    a. Consider any 10 triples from the Netflix RDF graph where some of the objects are literals and others are IRIs. Represent the information captured in these triples in the form of a property graph, where edges and nodes may have some properties associated with them. Draw the RDF graph and the equivalent Property Graph. **10 pt**
    b. Compare the indexing structures and the query optimization techniques used by any of the two triple stores (several triple stores are listed in the "Triple Stores" slides). Keep the discussion as technical as possible and the description should be of at most 1.5 pages in length. In the discussion, you can include the pros and cons of the indexing structures and the query optimization used in the respective stores. **10 pt**
    c. Compare the indexing structures and the query optimization techniques of any triple store with that of any property graph store. Keep the discussion as technical as possible and the description should be of at most 1.5 pages in length. Some of the popular property graph stores are Neo4j, JanusGraph, RedisGraph, Amazon Neptune, etc. In the discussion, you can include the pros and cons of the indexing structures and the query optimization used in the respective stores. **10 pt**