Semantic Web Team Projects

Spring 2016

The goal of the course project is to provide depth in a particular area of Semantic Web technologies in a hands-on fashion.

Overview

Your project must demonstrate **integration involving semantic web technologies**. The project is to use multiple sets of semantically enabled data (and generally already available).

Your team will select and perform one of the following:

1. LOGD Project: Use the framework and data provided at the LOGD (Linking Open Government Data) web site as the foundation for your project.
2. Custom Project: Select a framework and data for your project – preferably to solve an existing issue at one of the team member's place of work. The project must clearly be different than those performed with the LOGD framework for the extra value points.

The project requires understanding the semantic representation data sources and use of triples from multiple data sets to represent knowledge not present in any single set.

Here are some things to consider when defining your project:

- The project needs to result in a concrete outcome that will offer an interesting result or insight. Ultimately, the primary goal is that you learn something.

LOGD Projects shall be done in teams of 2 students. Custom Projects may have up to 4. Contact the instructor before the project proposal deadline if you want to propose a project with a different team size. NOTE: A larger team is potentially acceptable whereas a project involving a single individual is not.

The specific deliverables for the project include a proposal, one check point, and a final report as described in the following sections.

Project Proposal

The project proposal should include the following information:

- Team Name and Names of team members
- Type of Project (LOGD Project or Custom Project)
- Project Title and Topic
- Description of the specific data sources you intend to use in the project
    - LOGD Project – data source must originate from LOGD (http://logd.tw.rpi.edu/)
        - Identify how many triples are available at the LOGD SPARQL endpoint
        - It is very important to understand that the provided SPARQL endpoint

(http://logd.tw.rpi.edu/ws/sparqlproxy.php) does not have all the data shown converted to triples. You can determine the actual datasets available in the SPARQL endpoint by going to web site http://logd.tw.rpi.edu/sparql and issuing the default query provided there. For reference the query is:

```
PREFIX conversion: <http://purl.org/twc/vocab/conversion/>
SELECT ?g sum( ?triples ) as ?estimated_triples
WHERE { GRAPH ?g { ?g void:subset ?subdataset . ?subdataset conversion:num_triples ?
triples . } } GROUP BY ?g
```

- Custom Project
  - Describe the data to be used, including how many entities, how large, how or where is the data available?
  - What tools are needed for your project?
  - Describe your basic schedule for achieving the project with a few planned milestones where each milestone includes a date and planned progress at that point is generally sufficient.
    - There needs to be at least a milestone at or before the project checkpoint and plan to have your integration work done at least 1 week before the final report submittal is due.

- Expected Results - The knowledge/information to be presented by your visualization and why it is interesting (or at least who might find it interesting and why)
  - The project must be essentially different from any demonstration on LOGD or a project from another current or prior team from the Semantic Web class. NOTE: One different or additional significant and useful data source will suffice to make a project "essentially different" from another project.

Project proposals will typically be about a page in length. Project proposals must be submitted through eLearning. The project proposal is graded based on the above content and is a part of the project grade.

Project Checkpoint

There will be one project checkpoint. Checkpoints are short status reports that cover the following:

- Project Title
- Team Information (team name and members)
- Progress so far:
  - LOGD Project
    - LOGD Tutorials (what was completed, like...):
      - Building LOGD Visualizations
      - Understanding LOGD Data
      - Understanding LOGD Metadata
      - Mashing up LOGD data with SPARQL
      - How to find datasets using the LOGD SPARQL endpoint

        - …any other research you did
      - SPARQL query progress
    - Custom Project
      - LOGD Tutorials, if relevant to the custom project (as above)
      - Progress versus planned milestones
      - Progress at resolving open design issues
      - Changes in tools, tool use, and other tool issues including SPARQL
- Any results so far
- Problems encountered
- Changes in the project plan, like...
    - Change of data sources
    - Change in expected results

Provide some text in each category even if no changes were made. In other words, document things like:

- "No change in data sources" and
- "No change in the expected results"

to indicate you are maintaining

The project checkpoint will typically be about a page in length. The project checkpoint is graded based on the above content and is a part of the project grade.

Final Report

Each team must submit a final report. The final report includes both of the following:

- A written report
- A movie and/or interactive website of the project

The movie or web site is used to validate that actual software was written and functions. If a movie is used, audio is optional but preferred. Movies must be no more than 2 minutes in duration.

The written report is expected to require at least a couple of pages, but must be less than ten. There is a required outline, detailed below, for the written project. You may add subsections (i.e., 1.1, 1.2) as needed. The required outline of the written report is:

- Title
- Team Information (team name and members)
- Type of Project (LOGD Project or Custom Project)

1. Introduction
2. Target Audience – who the project is useful to and why

3. Description of Data Sources – ontology information is useful here, graphics are very helpful
4. Data Integration – what and/or how is the data being integrated (this may also include additional description of the value of the integration)
5. Data Product Results – text describing results with additional "value added" screen captures of your tool as needed
6. One of the following:
    - [LOGD Projects ONLY] Summary – summarize the previous sections highlighting critical information; include any outstanding issues, lessons learned, things that could have been done differently
    - [Custom Projects ONLY] Custom Project Justification and Summary – describe how this project is substantially different than the LOGD project with a comparative analysis using contrasting arguments.

Final Report Presentation

The team must present their final report. The last 4 sessions are reserved for team project presentations.

- Each team member should present equal portions of the final report.
- The presentation should highlight the project, its difficulties, and successes.
- The presentation should be no more than 15 minutes.

Milestones

End of Week 3: Submit project team information: 2 students for LOGD, up to 4 students for Custom

End of Week 8 (Oct 16): Project Proposals due

End of Week 11 (Nov 6): Project Checkpoint due

End of Week 13 (Nov 20): Final Report due

Week 14-15 (Nov 30, Dec 2, Dec 7, Dec 9: Final Report Presentations

Example Topics

See demonstrations on LOGD for examples. Your submission must be different from any example shown there. Use different data or valuable additional sets as needed.

Grading:

LOGD Project – Maximum: 100

Custom Projects – Maximum: 108

- Project Proposal: 20%
- Check point: 20%
- Final Reporting: 60% for LOGD,
  68% for Custom

Final Report Deductions:

The following elements are used as guidelines to deduct points from your final report.

Topic	**Points Deducted**
[Custom Project ONLY] A LOGD project by any other name	-9
Reuse of a prior semantic web project with no substantial change or extension	-60
Nothing submitted	-60 (LOGD) or -68 (Custom)
No introduction	-4
Weak target audience description	-7
Weak description of data sources	-6
Weak description of data integration	-7
Weak description of data product results	-8
Weak summary	-2
Written report is deficient, such as bullet points, overview, generalities versus detailed information using words and paragraphs	-10
Written report ignores formatting instructions	-10
Hard to understand the report throughout	-10
No movie or website (or website doesn't work)	-10
Movie longer than 2 minutes	-5 per minute, -10 max