COMP9322 Assignment 2: Semantic Modelling of an Information System

Due date: 1st of May 5 p.m. (Wednesday of Week 11)

General information

- This is an individual assignment
- You can ask questions and show your solution during the tutorials
- The deadline is 1st of May 5 p.m. (Wednesday of Week 11).
- You need to submit a zip file containing report and implementation, named as ***<zID>_<NameInWebCMS>_Assignment2.zip*** to the following Dropbox link https://www.dropbox.com/request/MDS01Q4ou2Vs1wzfLTdA

Business Case

Our business is predicting residential property price for corporate clients. We develop analytical models specifically tailored to the needs of our clients. Each analytical model is differentiated by:

- the training datasets that have been used for building the model
- the type of prediction model that have been used to create the model (e.g. regression, neural network etc.)

Customers can create new models by either supplying new training data (obtained from a data source) or changing the model type or both. They would like to keep access to previously used models just in case they want to revert to an old model, to reuse it or change it.

Information system requirements

The information system should allow the management of several analytical models associated with a number of customers.

Each model consists of:

- The model type: e.g. linear regression, neural network
- The training datasets used to build the model
- The code of the model (i.e. program/software module)

Training datasets can be obtained from a number of data sources using different software modules. The company also has a number of model building software modules implementing different model types. These software modules are upgraded/updated from time to time.

When the system becomes operational, any customer is able to:

- Select a data source and import datasets
- Create new models by choosing the appropriate model type and dataset

Example

The data sources in this case are:

- Domain: datasets can be downloaded by invoking "DownloadFromDomain" program
- Core Logic: datasets can be downloaded by invoking "DownloadFromCoreLogic" program

We assume we have 3 model building software modules:

- "Regression_Domain": this program takes a training dataset from Domain data source and creates a new model of regression type (e.g. "Predictor1_Reg_20190315").
- "Regression_CoreLogic": this program takes a training dataset from Core Logic data source and creates a new model of regression type (e.g. "Predictor2_Reg_20190315").
- "Neural_Network": this program takes a training dataset from either data source and creates a new model of neural network type (e.g. "Predictor_NN_20190315").

Deliverable

You need to create a **semantic information model** of the information system described above.

Your submission should contain:

- A report describing your model with diagram(s) and an example
- An implementation of the model and instance data (that corresponds to the example in your report) created using the Protégé tool and saved in Turtle format (.ttl)

Evaluation Rubric

- Identify and accurately modeling main classes, object properties and data properties. -30%
- Identify and accurately modeling main instance data. – 20%
- Introducing new classes and properties to enhance the capabilities of the ontology. – 10%
- Description of the model (in report) – 15%
- Justification of your modeling decisions (in report) – 10%
- Illustrate capabilities of your model through example queries, use cases and data (in report). – 15%