

This document describes how to use the alpha version of the RDR framework. Please note that since this is an early alpha release the software is released with no warranty, or any kind of guarantee against malfunction. There's going to be bugs, gentlemen.

INSTALLING AND LAUNCHING

Installing the software should be a simple process on Windows XP or newer, or Mac OSX 10+ operating systems, and should work without incident on most Linux operating systems, although this has not been tested. The only pre-requisite for the stand-alone version of this software is that the java runtime environment version 1.6 or newer be installed. The java runtime environment is available for free download at:

<http://www.java.com/en/download/index.jsp>

To run the RDR framework on most modern operating systems simply download the RDRFramework.jar file, place it in a folder of your choosing, and double click it.

Note: The RDR framework will create a number of files and folders when you begin using it, these will be placed in the same folder that the RDRFramework.jar resides in. As such, it is strongly recommended that you create a unique folder for RDRFramework.jar to reside in.

Alternatively, you may launch it from the command line. Assuming you have the java runtime environment correctly installed and configured, this can be done by opening a console, navigating to the folder which contains the RDRFramework.jar file, and typing:

```
Java -jar RDRFramework.jar
```

Note: If you receive an error which states that the "java" command cannot be found, your java runtime environment is not correctly configured. You may need to add the /bin folder of your java install to your operating system's PATH environment variable.

However, it is **strongly recommended** that you launch the program with additional heap memory, particularly if you intend to use simulated experts. To do this, instead launch the program with the command line:

```
Java -Xms128m -Xmx1024m -jar RDRFramework.jar
```

Note: It is recommended that you set up a launch script to do this for you. In Windows you can create a batch file, in Unix based operating systems you can create a shell script.

Once you have installed and launched the program, you will be presented with the main menu GUI shown in Figure 1.

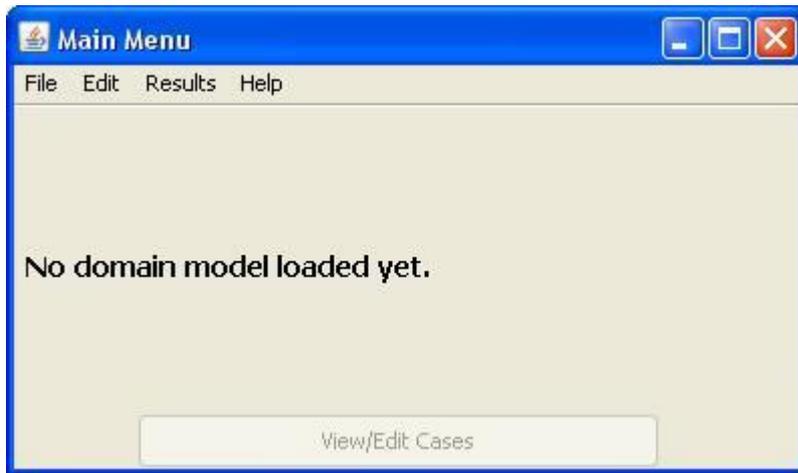


Figure 1 The main menu GUI.

Since you have not yet loaded a domain model, some of the options are greyed out. To access these, you will need to either: create a new domain model, load an existing domain model, or import a domain model from an *.arff* file.

CREATING/EDITING A DOMAIN MODEL

A *domain model* is a model which defines the structure of a domain. For example, a student record domain model might include a studentID (numeric), FirstName (nominal), LastName (nominal), Address (nominal), BirthDate (date), and Degree (nominal).

In the RDR framework software it is possible to define a domain model manually using the Domain Model Editor. This is useful when developing an experiment for a domain which has never been formally defined using either the RDR framework system, or for an *.arff* compatible piece of software, such as Weka. If you have previously defined a domain model, or have an existing *.arff* file for the domain you want to work with, move on to the next section.

To create a new domain model, click on the File menu, and select "New Domain Model", as shown in Figure 2.

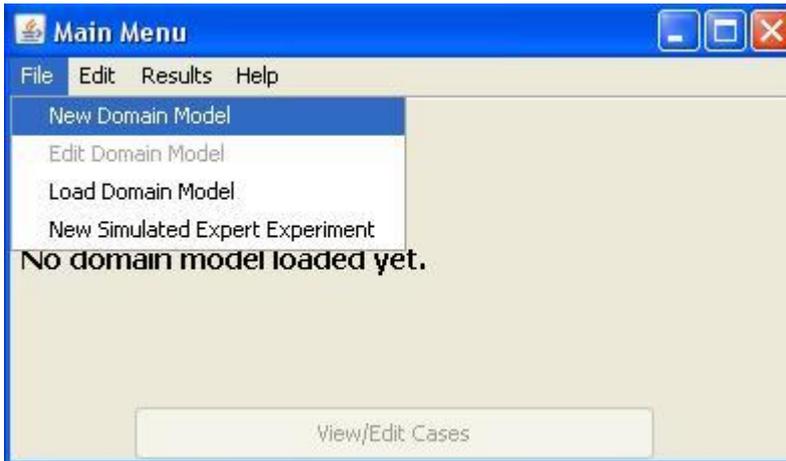


Figure 2 The file menu.

When this is done, the Domain Model Editor GUI will appear, with all fields empty, as shown in Figure 3.

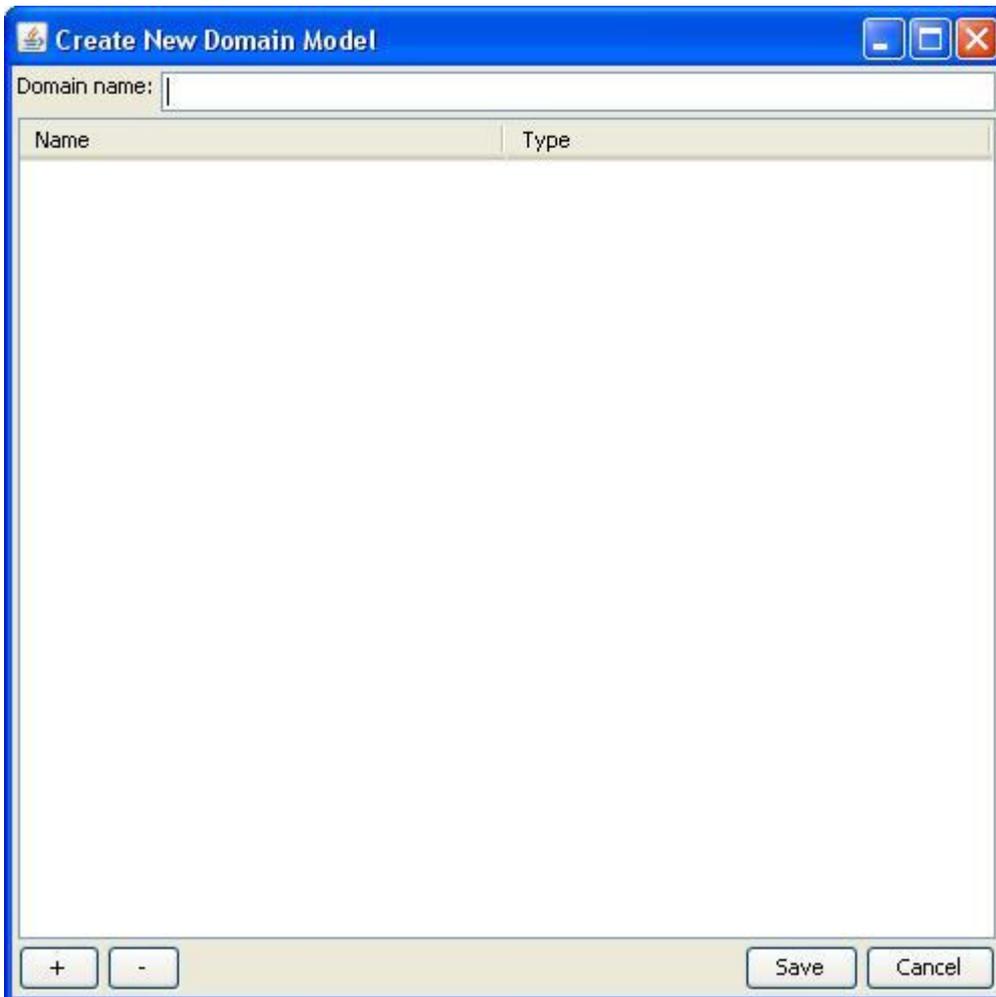


Figure 3 The Domain Model Editor GUI.

First it is important to name the domain model, by typing a name into the “Domain name” field. This name should not be the same as an existing domain name on your system.

To add attributes to the domain model, click on the “+” button on the bottom left. You may then name this attribute by double clicking on the left field in the table. You can select the attribute’s type by clicking on the right field in the table and selecting a type from the drop down menu which appears.

Note: Only the attribute types Nominal and Numeric have undergone any degree of testing at this stage. Other types may or may not work at this stage, and are likely to have bugs associated with them. Use with caution.

If we consider our earlier example of a student record, you may end up with a domain model definition such as the one shown in Figure 4.

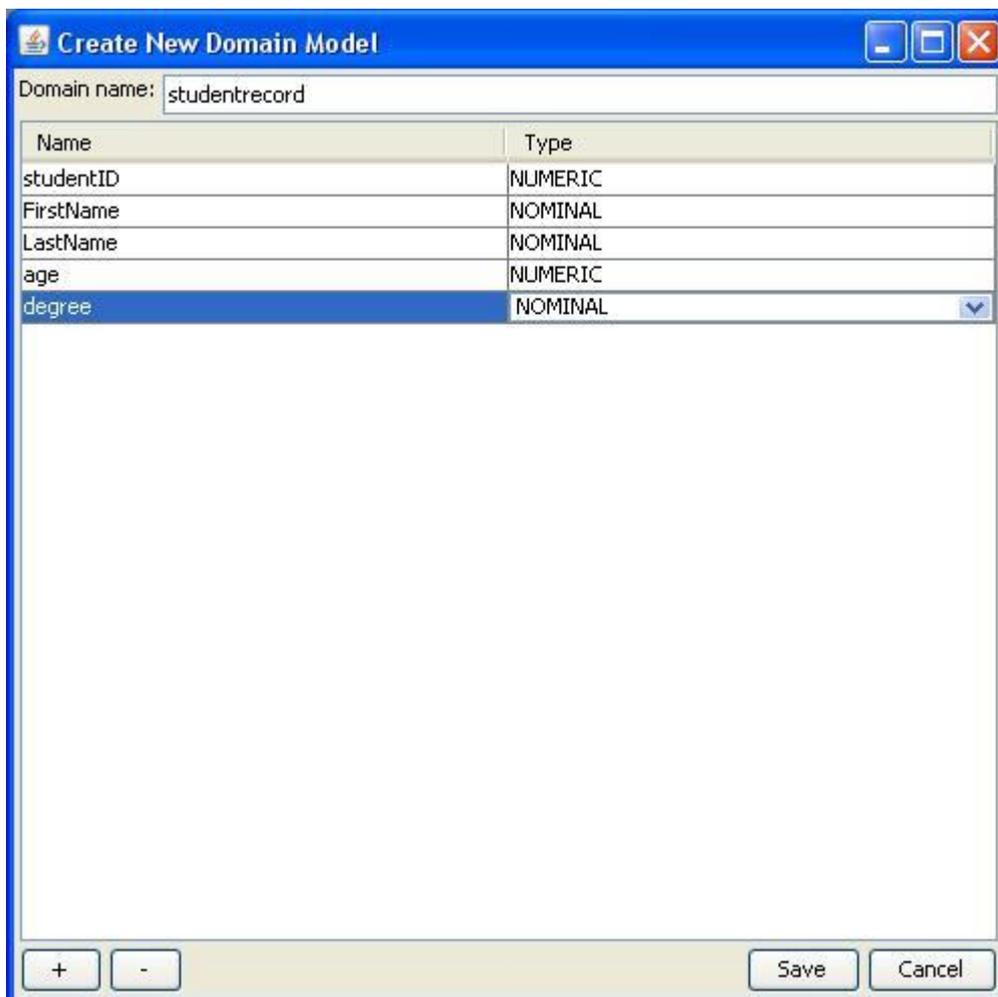


Figure 4 A sample domain model.

You may select an attribute and remove it from the model at any time by clicking the “-“ button. You may also cancel creating a domain model at any time by clicking the “Cancel” button.

Once you have defined your desired domain model, click the “Save” button to save it. This will send you back to the Main Menu GUI, but you will now be able to access the View/Edit Cases button, and the Edit Domain Model menu option.

To edit an existing domain model, select the Edit Domain Model option from the File menu. This is only possible once a domain model has been loaded. As long as the current domain has no cases defined, you may edit any aspect of the domain model. Once cases have been defined you will not be able to edit many parts of the domain model, as doing so would corrupt the existing cases.

LOAD/IMPORT AN EXISTING DOMAIN

You may load a pre-existing domain model file (.dm file) at any time using the “Load Domain Model” option in the File menu.

Clicking this option will bring up a file chooser dialog box, with which you can select either existing .dm domain model files, which are the RDR framework’s own domain model files, or you can select .arff files.

Once you have chosen a file, you will be taken back to the Main Menu, and once everything has loaded completely you will be able to access the View/Edit Cases button.

If you selected a .arff file then the program will initiate an import, converting the .arff data into the RDR framework’s native format, thus producing a .dm file and a .cas case file. During this import process you will be asked to identify how many classifications the .arff file contains. You should be able to find this out by consulting the source of the .arff file. You will also be asked if you wish to pre-populate the knowledge base using a simulated expert. If you click “Yes”, a simulated expert will run through the .arff file and produce an MCRDR knowledge base for you to work from.

Note: For large datasets the simulated expert process may take a (very) long time. Be warned, you cannot currently abort a simulated expert once it is commenced... you may need to terminate the program using your operating system if you are not prepared to wait for the process to complete. Also, for many datasets the default heap size will be insufficient to allow a simulated expert to run, see Installing and Launching for more information.

Once you have created or loaded a domain, you can access the Case Browser GUI by clicking on the View/Edit Cases button in the Main Menu.

If you have just created a new domain model, the case browser will show no cases, since you haven't added any yet. Otherwise, the list of existing cases will be presented in the table, as seen in Figure 5.

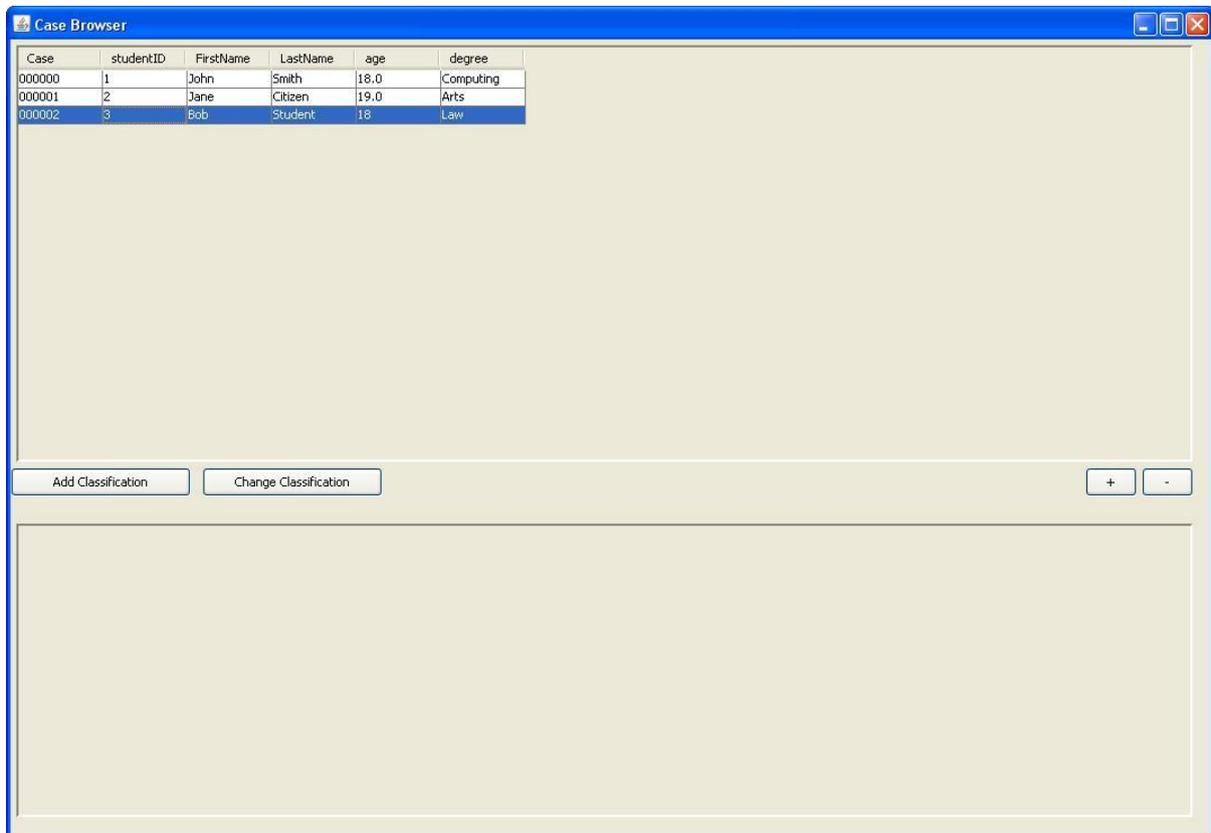


Figure 5 The Case Browser GUI when an MCRDR classifier is being used.

The case browser will always have the Change Classification, +, and – buttons visible. The Add Classification button will be visible for classifiers which can have multiple classifications.

You can click on any field in the case table to edit it, pressing Enter or clicking on another field will save the change.

If you click the + button on the middle-right, a new entry will be added to the case list, which can be edited in the same manner as above.

If you select a case, and click on the – button, the case will be deleted.

Clicking the Add or Change Classification buttons while a case is selected will trigger the Knowledge Acquisition GUI for the currently loaded classifier to launch, in which you can create a rule. If you currently have MCRDR selected as a classifier, the Change Classification button will only work if you have selected a classification from the lower panel of the GUI.

CREATING A RULE

Once you have selected a case and clicked the Add or Change classification button, the Knowledge Acquisition GUI will be displayed, as shown in . The operation of this GUI may change depending on the classifier you have loaded. The example shown here was using the MCRDR classifier on the studentrecord domain discussed earlier.

Attribute	Operator	Value
studentID	==	

Attribute	Value
studentID	1.0
FirstName	John
LastName	Smith
age	18.0
degree	Computing

Attribute	Value
-----------	-------

Figure 6 The Knowledge Acquisition GUI.

First, you should select what the classification of your rule will be. This can be done using the drop down box at the top, which lists all previously used classifications for this domain. Alternatively, you can add a new classification by entering the name into the text field and clicking the “Add” button.

Note: If the classification you add already exists in the classifications list it will not be added again, it will instead select the pre-existing classification from the list. This is to prevent duplicate classifications, however it also helps by acting as a search option in domains that have a very large number of classifications.

Having selected the classification, you must add the rule conditions. This can be done in one of two ways. For each condition in the rule, you may select the attribute, operator, and value manually. Alternatively, you can click on the attributes of the case shown in the table on the bottom left. Clicking once will pre-populate the drop down boxes and the value field with the values from the case. Double clicking will automatically add the condition to the rule. You may add as many conditions to the rule as you like.

You may select conditions you have previously added and remove them using the “Delete Selected” button at any time up until the rule is formally added.

Once you have added all the conditions to the rule, and selected a classification, you can attempt to Validate the rule by clicking Validate New Rule. If cornerstone cases are detected, one will be shown in the table on the bottom right. If there are multiple cornerstone cases detected, you can navigate between them using the < and > buttons.

In order for the rule to be added successfully, all detected cornerstone cases will need to be either eliminated or Permitted. You can permit a cornerstone case using the Permit button. To eliminate cornerstone cases, add additional conditions to your rule which are satisfied by the current case, but not by one or more cornerstone cases. When you add an additional condition you may click the Validate New Rule button again. Repeat this process until you have successfully added the rule.

To help in selecting differences, it is possible to select conditions from the cornerstone case by single or double clicking the appropriate field in the bottom-right table. These conditions will be added to the rule in much the same way as can be done using the current case table, but the default operator will be != instead of ==.

If a single classification classifier is loaded, the Permit, < and > buttons will not be available, as a maximum of one cornerstone case can exist for each rule, and a difference condition must be added to the rule before the rule can be successfully added.

CHANGING SETTINGS

To edit settings, select the Settings option from the Edit menu available on the Main Menu. This will launch the settings editor, as seen in Figure 7.



Figure 7 The settings editor.

In the current release, only two settings are available. Classifier, and Save Frequency.

Two classifiers are currently implemented, MCRDR and SCRDR. You may select either of these using the drop down menu.

The save frequency refers to how frequently the log file will be saved. i.e. if you set it to 10 then the log file will be saved every 10 entries. Setting this to -1 ensures that the log file will only be saved when the program is closed, the settings are changed, or a new domain

model is about to be created or loaded. It is recommended you keep the save frequency set to -1, unless you are worried about system instability and do not want to lose log records.

When you click the Save button the system will save the log file and classifier for the previously loaded classifier, and then load the log file and classifier for the classifier type you have selected.

PRODUCING EXPERIMENTAL RESULTS

If you have a domain and classifier loaded you can output experimental results at any time. To do this, select Output Basic Results from the Results menu available from the Main Menu.

This will output a set of comma separated value files to a location on your hard-drive. The location will be:

<current>/data/<domain>/results/<classifier type>/

Where current is the folder that the RDR framework is running from, domain is the name of the domain model you have loaded, and classifier type is the name of the classifier you currently have loaded (i.e. MCRDR or SCRDR).

If the output files already exist, the system will add the new results to the existing data. Newer data is always output on the far left set of columns. In this manner you can run an experiment on a domain several times, and assuming you keep the existing results files, each experimental outcome will be included in their own columns for easy comparison during data analysis. This is perhaps of most value when performing a simulated expert experiment, since each *fold* of the experiment will produce an additional column (or columns when appropriate) into each result file.

Note: Please ensure you do not have any of the results files open in other programs (e.g. Excel) when the system is attempting to write to them.

PERFORMING A SIMULATED EXPERT EXPERIMENT

To perform a simulated expert experiment, select File->New Simulated Expert Experiment from the Main Menu. This will launch the Experiment Builder GUI shown in Figure 8.

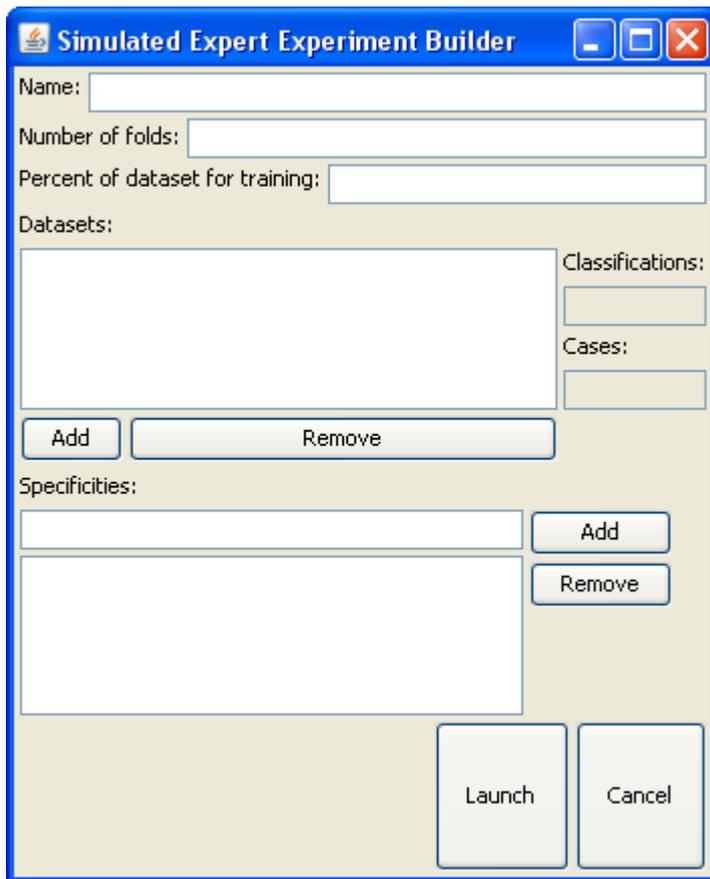


Figure 8 The Experiment Builder GUI.

To setup the simulated expert experiment, fill in all the fields then click the Launch button. Each of the fields are explained below:

Name: the name of the experiment. The outcomes of the experiment will be saved into a folder which matches this name. This name should be a unique name, that has not been used for an experiment previously.

Number of folds: the number of times the experiment will be repeated. It is important that this number is large enough to provide a good sample, despite the inherent randomness involved in the simulated expert's knowledge. It is recommended that for complex experiments, you use at least 10 folds.

Percent of dataset for training: determines how much of the datasets will be used to train the simulated expert. Should be a number between 0 and 1 i.e. 0.8 = 80%.

Datasets: you must add one or more datasets (.arff files) to the experiment. Each of the selected datasets will have a simulated expert trained for them, then use that simulated expert to produce an MCRDR knowledge base. For each added dataset, you must enter the number of Classifications and the number of Cases to the relevant fields. You can do this by selecting the added dataset from the list, then typing each number in to the appropriate field.

Specificities: the specificity is the proportion of the available conditions from the machine learned knowledge base that the simulated expert will attempt to use to create their own MCRDR rules. You can enter multiple specificities, which will result in multiple simulated experts for each fold for each dataset. Each specificity should be a number between 0 and 1, i.e. 0.5 = 50%.

When you click the Launch button, the simulated expert controller will take the parameters you have entered and commence the experiment. When the experiment is complete (which may take a very long time), the results will be output to:

<current>/data/<experiment name>/results/<dataset>/<classifier type>

If multiple datasets were selected, multiple output folders will be created, one for each dataset.

In each output file there will be columns corresponding to the number of folds and the number of specificities you selected. i.e. for a 3 fold experiment with 2 specificities 6 sets of output data will be created for each dataset. The specificity that the output corresponds to will be indicated in the column title in the .csv file.